Antonin MILZA
51 Bd Auguste Blanqui (Hall D) - 75013 Paris
06 32 90 85 76
antonin.milza@telecom-paristech.fr

Paris, le 02/29/2014

Objet : Note mettant en évidence le travail personnel le plus important réalisé pendant le cursus de formation à Telecom ParisTech

Monsieur le vice-président du Conseil général de l'économie, de l'industrie, de l'énergie et des technologies,

Je voudrais vous faire part de mon expérience dans le département de cybersécurité d'Airbus Defense&Space (Cassidian Cybersecurity). Ce stage ingénieur de six mois constitue à mes yeux le travail le plus abouti de ma formation et c'est ce qui me pousse à vouloir vous en parler davantage. Cette note est structurée en trois parties : une première partie introductive où j'explique ce qui m'a motivé à choisir la cybersécurité et plus particulièrement ce sujet de stage, une seconde partie relative à mon domaine de recherche et aux opportunités que j'ai identifié, et une troisième et dernière partie consacrée à mon prototype d'application et à sa valorisation.

Plan :

I. Introduction générale

Le secteur des industries de souveraineté et plus particulièrement la sécurité numérique m'attire depuis longtemps. J'ai été frappé ces dernières années par l'augmentation du nombre d'attaques informatiques, de plus en plus ciblées et sophistiquées, contre nos entreprises et administrations. L'attaque contre le ministère de l'Économie et des Finances lors du sommet du G20 de février 2011 en est un parfait exemple. Par ailleurs, lors de mes stages en laboratoire, PME et administration, j'ai remarqué que le chiffrement des données sensibles n'est aujourd'hui que très rarement utilisé. J'ai voulu comprendre pourquoi, face au danger des cyber-attaques, certaines de nos entreprises et organisations continuent à ne pas utiliser cet outil à leur disposition pour se protéger. Ces raisons m'ont incite à effectuer mon stage ingénieur dans le domaine de la cybersécurité et plus particulièrement chez Airbus Defense&Space.

Airbus Defense&Space, la division du groupe Airbus en charge des produits et services destinés aux usages militaires et astronautiques, a créé il y a 2 ans un département de cybersécurité composé d'une centaine d'experts dont la mission est de développer des solutions innovantes pour protéger les entreprises, gouvernements et infrastructures critiques contre le cyber-espionnage et le cyber-sabotage (anciennement Cassidian Cybersecurity). Le sujet de stage qui m'a été proposé porté sur l'exploration d'un domaine de la cryptographie moderne, encore cantonné au monde de la recherche académique, appelée Cryptographie Basée sur l'Identité (ou IBC pour *Indentity Based Cryptography*). Ce domaine, que j'avais repéré lors de mes lectures, me semblait extrêmement prometteur et j'avais l'intuition qu'il pourrait

m'aider à répondre à mon interrogation sur la faible utilisation du chiffrement des données. Ma mission a consisté à faire un état de l'art du domaine, à en identifier les opportunités industrielles et les contraintes sous la forme d'un livre blanc, et pourquoi pas à aller plus loin en développant un prototype d'application.

Mon étude a montré qu'il existe aujourd'hui deux mouvements antagonistes. D'une part la lutte pour la sécurisation des données dans un contexte d'espionnage mondial, et d'autre part la volonté de pouvoir s'échanger toujours plus facilement et à moindre frais les données entre individus et organisations. Mon étude, ainsi que le prototype que j'ai développé, tendent à montrer que la Cryptographie Basée sur l'Identité constitue un début de réponse permettant de réconcilier ces deux enjeux apparemment contradictoires. La suite de cette note vous permettra de mieux comprendre le domaine de recherche que j'ai étudié ainsi que les résultats que j'ai obtenu.

II. Etude du domaine de recherche et identification des opportunités

a) Avant-propos

La Cryptographie Basée sur l'Identité a été imaginée en 1984 par Adi Shamir, célèbre mathématicien et cryptographe à l'origine de l'algorithme RSA lors d'un exercice de pensée. L'IBC ne restera qu'au stade de l'idée abstraite jusqu'en 2001, lorsque deux chercheurs de l'université de Stanford proposèrent un algorithme de chiffrement basée sur l'identité (*Identity-Based Encryption from the Weil Pairing* Franklin, D. Boneh (1)). Entre 2001 et aujourd'hui des centaines d'articles scientifiques ont été publiés décrivant des protocoles cryptographiques basés sur l'identité toujours plus sophistiqués et sécurisés. A rebours, très peu d'applications pratiques de ces nouveaux algorithmes ont vu le jour. Cette forte différence entre monde académique et monde industriel s'explique entre autres, par la difficulté pour les développeurs du niveau applicatif de s'approprier les concepts mathématiques sous-jacents (en particulier les couplages, objet aux propriétés remarquables que je présenterai dans la suite de cette note) afin d'implémenter les algorithmes de l'IBC. En particulier, le choix du couplage, essentiel pour les performances et la sécurité du système, relève d'un équilibre subtil nécessitant une expertise poussée pour les courbes elliptiques.

Ainsi, la véritable motivation de mon travail pendant six mois a été de passer le cap théorique, en montrant que l'IBC, et plus particulièrement le chiffrement basé sur l'identité, pouvait être utilisé avec grand profit dans des applications pratiques au niveau industriel. La suite de cette note décrit les opportunités majeures de ce domaine que j'ai identifiées lors de mon stage, ainsi que le prototype d'application que j'ai développé.

b) La Cryptographie asymétrique traditionnelle

La cryptographie – discipline s'attachant à protéger des messages en s'aidant de secrets ou clés – est composée de différentes primitives. Dans cette note je me concentrerai sur la primitive de chiffrement, c'est-à-dire la transformation à l'aide d'une clé d'un message en clair (dit message clair) en un message incompréhensible (dit message chiffré) pour celui qui ne dispose pas de la clé de déchiffrement. Aujourd'hui, la plupart des communications de données chiffrées dans le monde utilise la cryptographie asymétrique et en particulier l'algorithme de chiffrement RSA. En permettant à deux individus qui ne se connaissent pas d'échanger des messages chiffrés, la cryptographie asymétrique a entrainé une révolution des usages depuis les années 1980 : aujourd'hui l'ensemble des échanges marchants effectués sur internet l'utilise. Néanmoins, nous allons voir que celle-ci n'est pas exempte de défaut, en particulier du fait d'une complexité certaine.

Avec la cryptographie asymétrique le chiffrement d'un message clair se fait à l'aide d'une clé publique (*i.e.* connu de tous) tandis que le déchiffrement d'un message chiffré se fait avec une clé privée (connu uniquement du destinataire du message). Ces deux clés (publique et privée) forment un couple et sont générées par chaque utilisateur. Précisons qu'une clé privée permet uniquement de déchiffrer les messages ayant étant chiffrés avec la clé publique du même couple car les celles-ci sont liées par construction algorithmique.

La cryptographie asymétrique repose sur une infrastructure à clé publique (ou PKI de l'anglais « *Public Key Infrastructure* ») qui a pour rôle de certifier les clés publiques des utilisateurs c'est-à-dire d'assurer à l'expéditeur que telle clé publique correspond bien à tel utilisateur. On dit que la PKI signe une clé publique avec un certificat. La tierce partie que représente la PKI a un rôle essentiel car elle permet de se prémunir d'attaques informatiques critiques appelées « attaque de l'homme au milieu » (de l'anglais « *Man in the Middle Attack* »). Ces attaquent permettent à un utilisateur malveillant de déchiffrer un message qui ne lui ai pas destiné en faisant passer sa clé publique pour celle du destinataire (Bob) auprès de l'expéditeur (Alice).

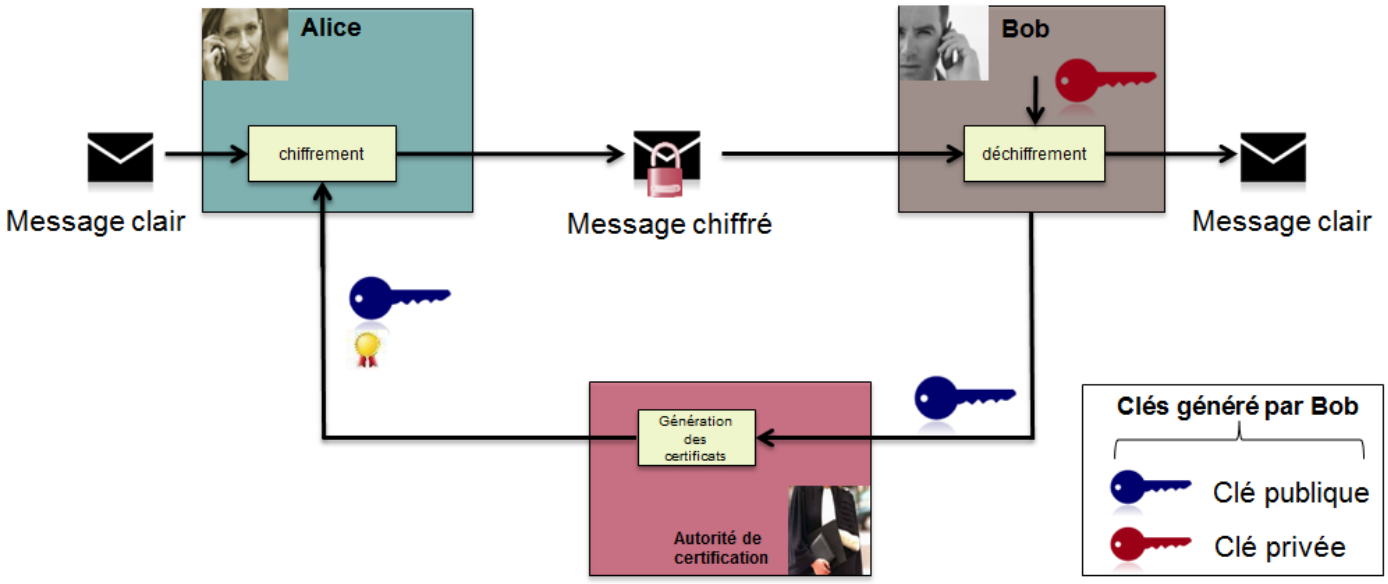

Figure 1 : Chiffrement et déchiffrement d'un message avec la cryptographie asymétrique basée sur une PKI.

C'est cette chaine de certification qui entraine une importante lourdeur des systèmes cryptographiques autant pour les utilisateurs finaux que pour les administrateurs. C'est le prix à payer pour mettre en échec les attaques précédentes. En particulier :

- Il faut administrer la PKI pour générer, mettre à jour et révoquer les certificats.
- Un tel système requiert d'importantes bases de données pour stocker les clés publiques certifiées.
- Un utilisateur souhaitant recevoir des messages chiffrés doit préalablement générer un couple de clés privée/publique, faire certifier sa clé publique par une PKI, et enfin la rendre accessible aux expéditeur potentiels.
- Un utilisateur ne peut envoyer des données chiffrées qu'aux utilisateurs ayant déjà une clé publique certifiée, et le cas échéant la récupérer.

Les systèmes reposant sur une PKI ont un double défaut : ils changent les habitudes des utilisateurs et demandent une administration du système particulièrement exigeante pour les entreprises. Ces défauts augmentent considérablement la barrière à l'entrée que doivent franchir les organisations pour installer, utiliser et administrer un système de chiffrement des données. Cette barrière à l'entrée est vraisemblablement trop haute aujourd'hui pour la plupart des entreprises et leurs employés puisque des études montrent que seulement 30% des entreprises dans le monde chiffrent leurs données. Les PME/PMI étant les plus en retard.

c) La Cryptographie Basée sur l'Identité (IBC)

Tout comme le chiffrement asymétrique traditionnel, le chiffrement basé sur l'identité repose sur l'existence d'une tierce partie, mais dans une version bien moins complexe que la PKI. Cette troisième entité est nommée *générateur de clés privées* (ou PKG de l'anglais « *Private Key Generator* ») car avec l'IBC ce ne sont plus les utilisateurs qui génèrent leurs clés privées mais cette nouvelle entité (S_{ID} sur la figure 2).

La cryptographie basée sur l'identité (IBC) est un type de processus cryptographique asymétrique dans lequel une chaine de caractères arbitraire est utilisée comme clé publique. Historiquement cette chaine de caractère représentait l'identité du destinataire. Par identité on entend toute chaine de caractère qui permet d'identifier un utilisateur de manière unique, par exemple une adresse email ou un matricule. Si l'on arrive à créer des clés de chiffrement secrètes relatives à ces identités de telle sorte que deux individus différents ne puissent avoir la même clé secrète, alors il n'est plus utile de certifier les clefs publiques, et la chaine de certification détaillée dans la partie précédente disparait.

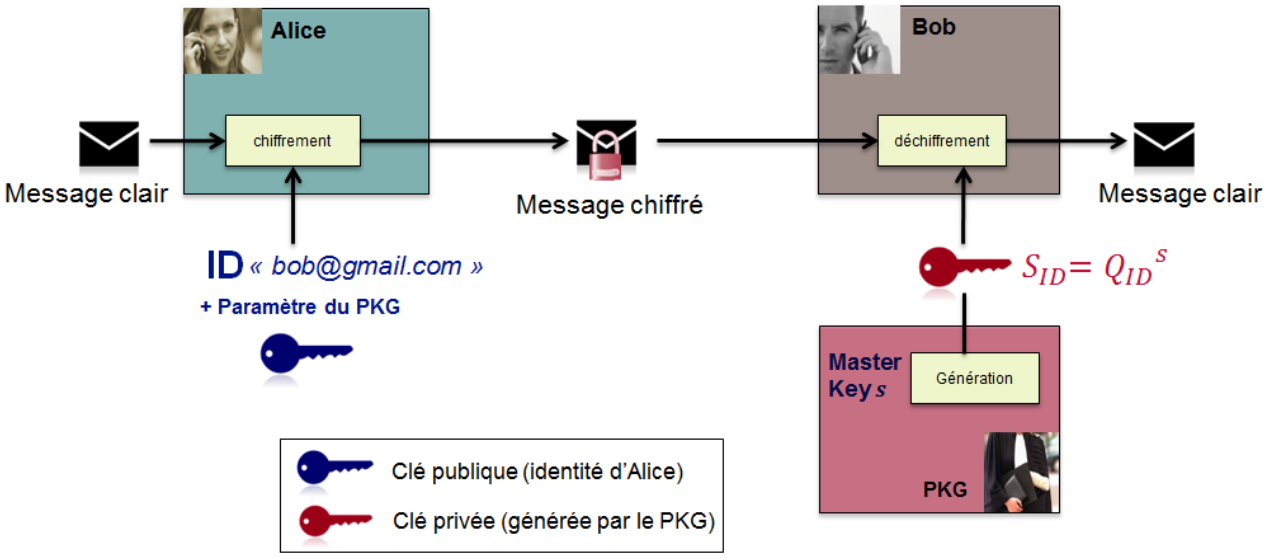

Figure 2 : Chiffrement et déchiffrement d'un message avec la cryptographie basée sur l'identité.

L'opportunité majeure de la cryptographie basée sur l'identité m'a ainsi semblé de simplifier fortement les systèmes cryptographiques pour les entreprises. Elle permet de réduire la « barrière à l'entrée » que doivent aujourd'hui franchir les organisations pour installer, utiliser et administrer un système de chiffrement des données car elle supprime une grande partie des défauts des systèmes reposant sur une PKI identifiée dans la partie précédente. En particulier :

- Plus de certificat, plus d'autorité de certification, plus de liste de révocation : on utilise une clé publique liée directement à l'identité des utilisateurs (une adresse email sur la figure 2).
- Plus de base de données pour stocker les clés publiques certifiées.
- Un utilisateur n'a plus aucune action à effectuer pour recevoir des données chiffrées.
- Un utilisateur souhaitant chiffrer des données n'a plus besoin de vérifier l'existence d'une clé publique certifiée, il utilise une chaine de caractères correspondant à l'identité de son destinataire.

Parmi les différents avantages de l'IBC c'est sa simplicité qui m'a le plus fortement intéressé lors de mon stage. Pour que la cryptographie et plus particulièrement le chiffrement des données soient massivement utilisés par nos entreprises, la simplification des systèmes est primordiale. Mes recherches et mon étude m'ont montré que l'IBC constituait un début de réponse, raison pour laquelle j'ai orienté la suite de mon stage dans cette direction.

d) Le fonctionnement mathématique de l'IBC

La complexité algorithmique de l'IBC vient du fait qu'il a fallu trouver un moyen de créer des couples de clés privée/publique qui soient liées entre elles par une même chaine de caractères (l'identité des utilisateurs) alors que ces deux clés sont ne sont générées ni aux mêmes moments ni par la même personne. En effet la clé publique est générée directement par l'expéditeur au moment du chiffrement en utilisant l'identité du destinataire alors que la clé privée (permettant le déchiffrement) est générée par le PKG sur demande du destinataire (la première fois seulement car le destinataire pour réutiliser cette même clé pour tous les messages chiffrés suivants). Le schéma théorique de l'IBC imaginé par Shamir en 1984 ne se concrétisera que vingt plus tard du fait de cette complexité. L'algorithme de chiffrement basé sur l'identité découvert en 2001 repose sur un objet mathématique bien particulier qu'on appelle couplage possédant la propriété de bilinéarité.

> **Couplage :** Soit G_1, G_2 et G_t trois groupes cycliques de même ordre q. Une application bilinéaire de $G_1 \times G_2$ vers G_t est une fonction $e : G_1 \times G_2 \rightarrow G_t$ telle que pour tous $u \in G_1$, $v \in G_2$ et $(a, b) \in \mathbb{Z}^2$:
>
> $$e(u \circ a, v \circ b) = e(u \circ b, v \circ a) = e(u, v) \odot a \odot b$$
>
> L'opérateur $\circ$ est une multiplication sur une courbe elliptique tandis que l'opérateur $\odot$ est une multiplication sur un corps fini.

C'est cette propriété qui permet de lier la clé publique et la clé privée de manière à ce que seule la clé privée correspondant à une clé publique donnée (et donc à une identité donnée) puisse déchiffrer le message. Nous montrons ci-dessous que grâce aux couplages et à leur propriété de bilinéarité le destinataire est capable de déchiffrer le message et qu'il est le seul à pouvoir le faire. Ci-dessous x désigne le résultat d'un couplage qui est calculé et utilisé par l'expéditeur pour chiffrer son message. On admettra que si le destinataire est capable de retrouver x, alors il est capable de déchiffrer le message.

$$\boldsymbol{x = e(Q_{ID}, P \circ s) \odot r}$$

- Q_{ID} est l'identité du destinataire (bob@gmail.con sur la figure 2).
- $P \circ s$ est un nombre généré par le PKG mais connu de tous (rendu public) qu'on appelle les paramètres du PKG.
- La variable s correspond à la « master private key » connu uniquement du PKG et lui permettant de générer la clé privée $S_{ID} = Q_{ID} \circ s$. Seul le PKG peut calculer S_{ID} (car seul lui connait la « master private key » s) et la transmettre à l'unique utilisateur légitime (c'est-à-dire avec la bonne adresse email).
- r est un nombre aléatoire généré par l'expéditeur et connu uniquement de lui.

On peut déjà remarquer que puisque l'expéditeur connait Q_{ID} (bob@gmail.con), $P \circ s$ (connu de tous) et r (généré par lui-même) il est capable de chiffrer le message (on ne détaille pas ici comment se fait le chiffrement à partir de x). De son coté, après réception du message chiffré, le destinataire va calculer le couplage suivant :

$$\boldsymbol{e(S_{ID}, P \circ r)}.$$

Le nombre $P \circ r$ est envoyé par l'expéditeur au destinataire avec le message chiffré, il est donc connu de ce dernier. Par ailleurs, S_{ID} correspond à la clé privée que le PKG a générée en utilisant l'identité du destinataire Q_{ID} (bob@gmail.con sur la figure 2). Puisque le destinataire (Bob) est bien légitime (il possède l'identité bob@gmail.con) il peut donc récupérer auprès du PKG sa clé privée S_{ID}. Ainsi, le destinataire peut calculer $e(S_{ID}, P \circ r)$ et il est le seul. On montre finalement que le fait de pouvoir calculer $e(S_{ID}, P \circ r)$ revient à connaitre x (et donc à déchiffrer le message) grâce à la propriété de bilinéarité :

$$\boldsymbol{e(S_{ID}, P \circ r) = e(Q_{ID} \circ s, P \circ r) = e(Q_{ID} \circ r, P \circ s) = e(Q_{ID}, P \circ s) \odot r = x}$$

$S_{ID} = Q_{ID} \circ s$ par definition

Bilinéarité du couplage

III. Prototype d'application et valorisation

Aujourd'hui en moyenne, 72 courriels sont reçus et 33 sont envoyés par jour et par employé. Ces échanges d'informations, outre le fait de croitre d'année en année, prennent aujourd'hui des formes différentes à travers des espaces collaboratifs et de partage tels ceux liés au « *cloud computing* ». Ces nouveaux moyens de communications sont très pratiques mais ils présentent aussi des risques majeurs du point de vue de la sécurité. Du fait de l'interconnexion de la plupart des entreprises au réseau Internet, le cyberespace est de plus en plus utilisé à des fins d'espionnages. Les attaques informatiques, hier opportunistes et diffuses, aujourd'hui ciblées et persistantes sont désormais motivées le plus souvent par l'obtention d'informations confidentielles.

J'ai donc voulut développer un prototype d'application répondant à la problématique de confidentialité des données dans un contexte de partage et de mise en commun de l'information tout en garantissant une sécurité optimale. J'ai repris les conclusions de mes recherches pour implémenter les algorithmes d'IBC et développer un prototype d'application utilisable et administrable par la plupart des entreprises. Mon prototype composé de diffèrent modules indépendants a été développé en langage C pour l'implémentation des algorithmes bas niveau et en C# pour l'interface utilisateur et son intégration à des produits existants. La problématique d'utilisabilité est à mon sens primordial pour généraliser l'utilisation du chiffrement des données. J'ai voulu proposer une version de mon prototype pour le logiciel Microsoft Outlook car celui-ci est largement déployé dans les entreprises et administrations pour l'envoi et la réception des emails. La première version de mon prototype permettait ainsi à un utilisateur de pouvoir chiffrer un email avec l'adresse email du ou des destinataire(s) en cliquant simplement sur un bouton dans l'interface utilisateur d'Outlook. Plus aucune action n'était alors nécessaire à l'utilisateur pour envoyer ou recevoir des emails chiffrés.

Fort de ce premier succès, j'ai souhaité aller plus loin. Mon objectif a été de rendre l'utilisation du chiffrement toujours plus simple et pratique pour les entreprises et les organisations. Aujourd'hui, une demande forte des entreprises est de pouvoir définir des contrôles d'accès aux données en se basant sur le rôle ou la fonction des individus, et non plus uniquement sur leur identité. Par exemple, dans un ministère, un Directeur Général peut vouloir chiffrer un email qui ne puisse être lu que par les chefs de bureau. Dans une entreprise, un fichier Excel partagé sur Microsoft SharePoint ne doit pouvoir être lu ou modifié que par les membres du département marketing. J'ai alors eu l'idée d'améliorer la première version de mon prototype en permettant aux utilisateurs de chiffrer leurs données non plus avec une adresse email, mais avec une chaine de caractère représentant le rôle du ou des destinataire(s). Lors de mon stage, j'ai particulièrement travaillé sur la problématique de la confidentialité des données personnelles de santé. Alors que le gouvernement souhaite relancer ce chantier important, à travers la généralisation du Dossier Médical Personnel (DMP), j'ai estimé qu'il s'agissait d'un cas d'usage intéressant pour mon prototype. Avec la nouvelle version de mon prototype, un médecin urgentiste pourrait chiffrer la radiographie d'un patient avec une clé publique « chirurgien » pour qu'elle ne soit consultable que par un chirurgien. La vérification qu'un individu possède bien le rôle adéquat dans l'hôpital est effectuée par une version modifiée du PKG que j'ai développé.

A la fin du développement, nous avons voulu avec mon maitre de stage généraliser l'utilisation de mon prototype. Deux projets ont été particulièrement envisagés : Arkoon, la filiale du groupe Airbus en charge des solutions de chiffrement réfléchit à une intégration de mon prototype pour la protection des emails et des services de *cloud computing* dans son logiciel SecureBox. Le programme de recherche et d'innovation de l'union européenne (Horizon 2000) qui s'intéresse à la problématique de la protection des *Smart Grids*, a lancé un appel d'offre sur ce sujet. Mon maitre de stage réfléchit à une réponse reprenant le concept de mon prototype. Enfin, dans mon souci de rendre le chiffrement des données toujours plus simple d'utilisation, j'ai eu l'idée d'utiliser mes connaissances du Big Data pour imaginer une solution de chiffrement automatique des emails en fonction de leur niveau de confidentialité. L'idée est d'analyser par *machine learning* les données à transmettre pour en déterminer un niveau de criticité. A partir d'un certain seuil, les données seront chiffrées automatiquement en utilisant l'IBC. Cette idée s'est concrétisée par une demande de dépôt de brevet au nom de mon maitre de stage et de moi-même, que j'ai rédigée avec le département de propriété intellectuelle d'Airbus *Defense and Space*. Nous y travaillons encore actuellement pour avoir toutes nos chances d'aboutir au brevet.