Cyber FizzBuzz : réflexion sur la sélection de candidats dans le domaine de la SSI

Éric Jaeger Olivier Levillain Christian Lixi

Résumé Pour le recrutement, il est utile de disposer de questions standards permettant d'apprécier rapidement l'aptitude des candidats à appréhender divers sujets et approches. Nous présentons dans cet article le processus de sélection des candidats à une formation technique en SSI – des questions parfois éloignées du cœur du sujet, mais dont nous justifions la pertinence.

1 Introduction

FizzBuzz

Jeff Atwood a publié sur son blog *Coding Horror* plusieurs articles sur le recrutement de développeurs, et décrit une question inspirée par le jeu enfantin FizzBuzz [1] :

> Write a program that prints the numbers from 1 to 100. But for multiples of three print "Fizz" instead of the number and for the multiples of five print "Buzz". For numbers which are multiples of both three and five print "FizzBuzz".

Bien entendu, il ne suffit pas d'être capable de répondre à cette question pour convaincre. Mais la réciproque est vraie : il n'est pas envisageable de recruter un développeur qui serait incapable de résoudre ce problème en quelques minutes. Or, plusieurs témoignages de recruteurs ayant tenté l'expérience attestent que, de manière assez surprenante, la majorité des candidats à un poste de développeur échouent à ce test.

La problématique existe dans tous les domaines, et en particulier en SSI : comment évaluer efficacement un candidat lors d'un entretien ? Après plusieurs années de recrutement pour des postes en laboratoire et de sélection de candidats à des formations, nous avons élaboré un jeu de questions, notre *Cyber Fizzbuzz*.

Vous avez dit sécurité ?

Qu'il s'agisse de définir des programmes de formation, d'élaborer des cours ou de collaborer avec des professionels de l'informatique non spécialisés en sécurité, il est souvent nécessaire de fournir des illustrations simples de ce qui distingue l'informatique classique de la SSI, parfois au travers de questions inhabituelles et inattendues.

Ce fut par exemple le cas lors des études sur la sécurité des langages de programmation (JavaSec [2] et LaFoSec [3]). Nous avons en effet constaté que les premiers mois des travaux servaient essentiellement à clarifier avec nos interlocuteur le sujet de l'étude... le temps nécessaire pour passer de l'analyse de performance des bibliothèques cryptographiques qui nous était initialement proposée, à une compréhension de l'intérêt (et de la faisabilité) pour un *garbage collector* d'effacer par surcharge les zones mémoires libérées.

1. `http://blog.codinghorror.com/why-cant-programmers-program/`
2. `https://www.ssi.gouv.fr/javasec`
3. `https://www.ssi.gouv.fr/lafosec`

La formation ESSI

La formation ESSI est un cursus de 13 mois qui mène à un titre RNCP de niveau I (équivalent Bac+5). Elle se compose de deux parties :

- une période scolaire de 7 mois (700 heures de cours/TP) ;
- un stage de 6 mois avec rapport et soutenance sur des sujets tels que les télévisions connectées, l'analyse d'un algorithme de chiffrement ou encore l'étude des formats de fichiers d'images.

Il s'agit d'une formation dense qui propose une vision technique et large de la sécurité. Elle a pour principes la curiosité et la rigueur. L'objectif est de former des personnes crédibles auprès des équipes techniques, et convaincantes auprès des décideurs.

En effet, la SSI ne se résume pas à la cryptographie – ni d'ailleurs au déploiement d'antivirus, aux normes ISO 27000, ou à la mise en place d'un *Web Application Firewall* réputé protéger contre les attaques du top 10 Owasp.

La SSI, pour un spécialiste technique, c'est aussi – voire surtout – se poser des tas de questions sur un système dans un contexte donné. Il s'agit par exemple de se demander comment se comporte un *parser* lorsqu'il traite un fichier non conforme au format attendu. Ou, constatant que l'adresse source d'un paquet IP n'étant pas utile pour son acheminement, de supposer qu'elle n'est pas vérifiée par les équipements de routage et peut donc être trafiquée à l'émission. Et ainsi de suite, sans se laisser arrêter par des considérations telles que « C'est conforme à la spécification, quel est le problème ? », « Mais non, cela ne peut pas fonctionner comme cela, ce n'est pas prévu pour. » ou encore la réplique très classique « Mais pourquoi quelqu'un ferait ça ? » – sans parler des *famous last words* « Il y a forcément quelqu'un qui a vérifié ! »

Cyber FizzBuzz

Nous nous concentrons plus spécifiquement dans la suite sur notre expérience concernant la sélection des candidats à la formation ESSI (expert en sécurité des systèmes d'information, voir encadré) dispensée au centre de formation de l'ANSSI. La sélection repose en particulier sur un questionnaire envoyé aux candidats en amont de l'entretien de sélection.

La particularité de ce questionnaire est qu'il vise à évaluer l'aptitude des candidats à appréhender les concepts utiles à la SSI, et non leurs connaissances. Il s'agit tout d'abord d'éviter l'accusation de mener des « tests Google »[4], qui n'auraient que peu d'intérêt en entrée de formation. Mais surtout, dans notre expérience, il est plus aisé pour les candidats ayant un certain état d'esprit d'acquérir les connaissances leur manquant que de faire le chemin inverse. Comme nous le verrons dans les exemples, il n'est pas nécessaire pour évaluer cet état d'esprit que la question traite explicitement de SSI.

2 Questions... et discussions

Abordons maintenant le questionnaire de sélection des candidats à la formation « Expert en SSI » du centre de formation de l'ANSSI.

Ce questionnaire doit être rempli à domicile, préalablement à l'entretien, en temps borné et sans aucune assistance (livre, aide d'un collègue, calculatrice ou ordinateur). Les notations

4. `http://zythom.blogspot.fr/2013/08/lanssi-et-le-test-google.html`

et les définitions utiles en mathématiques sont rappelées, pour les candidats qui pensent à consulter le mode d'emploi fourni en début questionnaire.

Pour l'essentiel, ce sont des questions à choix multiples – avec parfois des possibilités absurdes, par exemple indiquer qu'un « groupe » en algèbre est « carré », sans que cette notion ne soit définie. Ce type de proposition permet d'avoir des discussions intéressantes avec les candidats qui répondent à l'intuition (ou au hasard).

Ce questionnaire n'est pas noté ; c'est un outil permettant d'avoir un entretien efficace. L'absence de réponse à certaines question est tout à fait tolérée. Par contre, il est apprécié que le candidat, confronté à des sujets qu'il ne connaissait pas lorsqu'il a abordé le questionnaire, ait pris le temps de se renseigner avant l'entretien. Encore une fois il est essentiel de comprendre qu'au-delà des réponses apportées, ce qui nous intéresse est aussi la démarche et la justification apportée par le candidat.

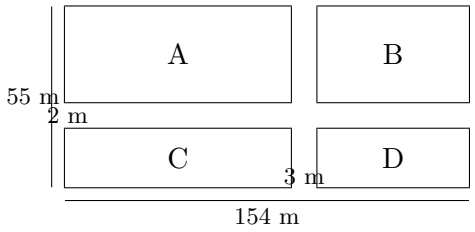
Un peu de mathématiques

Q Soient X un ensemble dont aucun élément n'est négatif et Y un ensemble dont aucun élément n'admet de racine carrée réelle. Que dire de l'affirmation « Tout réel n'appartenant pas à X appartient à Y » ?

La question ne fait que très modérément appel à des concepts mathématiques – pas de quoi déstabiliser un candidat appelé ensuite à interagir avec des cryptographes.

Elle permet surtout de repérer une tendance à généraliser un peu trop vite. Pour de nombreux candidats, si aucun élément de X n'est négatif, c'est donc qu'ils sont tous positifs ou nuls (ce qui est vrai)... et de conclure, à tort, que X est *égal* $\mathbb{R}^+$, et, de manière similaire, que l'ensemble Y est égal à $\mathbb{R}^{-\star}$. C'est cette généralisation abusive sur laquelle il faut attirer l'attention : baser une analyse ou un raisonnement sur la première solution trouvée à un problème est dangereux.

Q Soit un jardin dont la disposition et les dimensions en mètres sont données ci-dessous :



À raison de 17 bulbes par m^2, pour fleurir les parcelles A (de 21 m par 85 m), B, C et D il faut :

- □ 134147 bulbes
- □ 136051 bulbes
- □ 138323 bulbes

Ce n'est pas la capacité du candidat à effectuer des calculs à la main qui est évaluée ici [5]. Il s'agit plutôt de l'inviter à envisager qu'il peut y avoir plusieurs approches pour répondre à la question. L'approche systématique consistant à faire les calculs, qui fonctionnera dans tous les cas, est ici inefficace. Une approche plus subtile est possible, et permet de répondre mentalement et en quelques secondes – le type de stratégie qui sera retenue par un attaquant qui n'a cure de rester sur les sentiers battus.

5. Il faut répondre au questionnaire sans ordinateur.

Un peu de culture informatique

Q L'architecture de Von Neumann c'est...

- ☐ la structuration des données en programmation orientée objet
- ☐ une architecture pour ordinateurs qui ne permet pas de distinguer les programmes des données
- ☐ une décomposition des systèmes d'exploitation sous la forme de couches normalisées
- ☐ une alternative à l'architecture Harvard

Le candidat sérieux qui ne connaît pas cette notion de culture générale en informatique au moment de remplir le questionnaire aura eu le soin de consulter internet avant l'entretien. Mais qu'a-t-il retenu et que peut-il en déduire ? Quelles sont les alternatives, et les conséquences en termes de SSI ? Pour les candidats les plus avertis, la discussion peut se poursuivre sur les mécanismes permettant de restreindre l'exécution de certaines zones en mémoire.

Q Que pouvez-vous dire du nombre en hexadécimal `0x1616160` ?

- ☐ c'est un nombre pair ;
- ☐ c'est un multiple de 3 ;
- ☐ c'est un multiple de 16 ;
- ☐ c'est un multiple de `0x16` ;
- ☐ c'est un multiple de `0x10101`.

C'est (vaguement) une question d'informatique, mais clairement pas de SSI. Il y a certes quelques connaissances à avoir pour y répondre, mais savoir ce qu'est l'hexadécimal est utile pour discuter avec des informaticiens qui l'utilisent par automatisme sans même envisager que leur interlocuteur puisse ne pas connaître cette représentation.

Il ne s'agit toujours pas de vérifier que le candidat est capable de faire à la main de longs calculs mais d'aborder les sujets suivants :

— Pourquoi l'hexadécimal est utilisé en informatique ?
— Que signifie pour un nombre en hexadécimal de se terminer par `0` ?
— Le critère de divisibilité par trois [6] s'applique-t-il à un nombre écrit en hexadécimal ?

La réponse à ce dernier point est oui, mais il est attendu d'un candidat qu'il se pose la question plutôt que d'opérer par automatisme, et idéalement, de savoir pourquoi la réponse est vraie en base 10 comme en base 16.

Q Le complément à deux c'est...

- ☐ l'instruction assembleur multipliant par 2
- ☐ une astuce de programmation pour décomposer un calcul sur de grands entiers en traitements parallèles
- ☐ une convention permettant de représenter les entiers relatifs
- ☐ une représentation non optimale pour un entier naturel

Tous les candidats ne connaissent pas la notion de complément à 2, c'est admis. D'autres candidats connaissent, répondent à la question, et sont même capables d'expliquer comment on calcule un complément à 2.

La discussion peut alors prendre un tour intéressant : pourquoi quelque chose d'aussi compliqué, alors qu'il aurait suffit d'utiliser le premier bit comme bit de signe ? En la matière, la curiosité est de loin préférable à l'acceptation sans discussion d'un état de fait. Cette discussion peut alors mener à d'autres considérations, par exemple les conséquences sur la sécurité d'une confusion entre une représentation signée et une représentation non signée.

6. Un nombre est divisible par 3 si la somme de ces chiffres est divisible par 3.

Q La représentation des nombres en virgule flottante...
- □ permet de représenter n'importe quel réel
- □ est de taille fixe
- □ contient une indication de la position de la virgule

Il est toujours intéressant de discuter avec un candidat ayant répondu vrai aux deux premières assertions. Pour les candidats les plus à l'aise, il peut être intéressant lorsque cette question est abordée de s'intéresser à la représentation en base 2 de 1/2 ou de 1/10, ou encore d'aborder la question de l'égalité entre deux flottants.

Ces conversations sur la représentation des valeurs dans la mémoire des ordinateurs mènent parfois à des questions des candidats pour le moins déroutantes, par exemple « Mais comment l'ordinateur peut-il savoir si le nombre en mémoire est en décimal ou en hexadécimal ? » ou encore « Faut-il mettre un espace après le premier bit pour savoir qu'il correspond au signe et non à la valeur ? ».

Un peu de cryptologie

Bien que le questionnaire ne cherche pas à évaluer les connaissances en SSI, il ne faut pas décevoir le candidat qui s'attend à être interrogé avec de « vraies » questions sur la sécurité.

La cryptologie peut amener des questions intéressantes. Tout candidat a en effet utilisé la cryptographie sur internet (pour gérer un compte bancaire ou faire des achats en ligne), et il devrait s'être posé quelques questions simples, puisqu'il s'intéresse à la SSI. Mais il n'est pas question de demander le schéma d'un tour d'AES ou des éléments de théorie des courbes elliptiques.

Q Un certificat électronique...
- □ doit rester secret pour jouer son rôle ;
- □ permet l'authentification ;
- □ contient la clé utilisée pour déchiffrer les communications ;
- □ peut contenir une clé publique ECDSA.

Il existe une quantité de personnes souhaitant faire carrière en SSI qui se révèlent malheureusement incapables de répondre de manière simple et rigoureuse à cette question. Souvent, il est utile de revenir aux fondamentaux et de demander ce qu'est au fond un certificat. Obtenir une définition est très difficile, même de la part de candidats ayant dans un poste précédent mis en œuvre une infrastructure de gestion de clés.

Pour les candidats les plus instruits sur ces sujets, la discussion la plus intéressante sera de savoir quelles sont les garanties apportées par un certificat auto-signé.

Q Une fonction de chiffrement E prend en entrée une clé k et un message clair, et retourne un chiffré. Pour remplir son rôle, E_k doit être...
- □ injective
- □ surjective
- □ déterministe

Ici encore, il y a un peu de notions mathématiques – et les définitions utiles sont rappelées en début de questionnaire. Ce qui est intéressant, c'est de vérifier si le candidat s'est approprié ces notions en en comprenant les conséquences pratiques.

Ainsi, si le chiffrement n'est pas injectif, c'est que deux messages peuvent avoir le même chiffré... ce qui rendrait le déchiffrement problématique. La surjectivité permet d'aborder le sujet de la malléabilité. Le déterminisme permet d'avoir des discussions encore plus intéressantes :

que peut-on gagner ou perdre à garantir que le même message aura toujours le même chiffré, ou à introduire de l'aléa dans l'algorithme ?

Nous avons déjà indiqué qu'il faut éviter de confondre SSI et cryptographie. De plus nous préférons mettre en avant dans ce questionnaire l'usage de la cryptographie plutôt que la théorie mathématique. C'est à dire comprendre les différences fondamentales entre cryptographie symétrique et cryptographie asymétrique en termes de gestion de clés ou encore l'intérêt pratique de la *forward secrecy* plutôt que les équations RSA ou la difficulté du problème du logarithme discret.

Un peu de *shell*

Q Quelles commandes *shell* peuvent mener (sans redirection) à la perte de données ?

- ☐ `ls`
- ☐ `rm`
- ☐ `mv`
- ☐ `cat`
- ☐ `cp`

Cette question est souvent intéressante. De nombreux candidats la considèrent comme purement technique, visant à vérifier la connaissance des commandes Unix. Bien entendu, ce n'est pas du tout l'objectif poursuivi.

En effet, la réponse la plus fréquente se limite à la seule commande `rm`. Ce faisant, le candidat répond non pas à la question posée, mais à une interprétation de cette question, que nous qualifions de *fonctionnelle* : « Comment feriez-vous pour détruire des données ? ».

L'entretien doit permettre au contraire de faire comprendre à ces candidats que la portée de la question est bien plus large. `rm` permet bien de détruire un fichier et les données qu'il contient [7]. Mais si l'objectif est de préserver des données, `rm` n'est pas la seule commande dangereuse.

Q Un utilisateur tape la ligne « `cat *` » dans un *shell*. Parmi les propositions suivantes, lesquelles sont vraies ?

- ☐ le répertoire courant peut contenir un fichier nommé `*` dont le contenu sera affiché
- ☐ le répertoire courant peut contenir un fichier nommé `-v` dont le contenu sera affiché
- ☐ le répertoire courant peut contenir un fichier nommé `> rights.acl` dont la commande mènera à l'écrasement du fichier
- ☐ le répertoire courant peut contenir un fichier nommé `; rm *`, et les fichiers du répertoire seront effacés

Ici encore, il s'agit d'une question à tiroirs. Il faut d'abord souvent passer par le stade du déni – certains candidats considérant impossible de donner de tels noms à des fichiers, forts parfois d'essais menés au travers d'une interface graphique. La discussion peut alors se poursuivre sur les restrictions imposées par une telle interface, et les risques qu'il y a à les méconnaître.

Il est également d'usage de donner en exercice au candidat la création d'un fichier nommé `*` ou `-x` depuis la ligne de commande à l'aide de la commande `touch` [8].

Ensuite, il faut faire comprendre l'intérêt en SSI de cette question, que des administrateurs Unix expérimentés peuvent ne jamais s'être posée. Enfin, cela permet d'aborder la notion d'injection, et de comprendre le niveau d'analyse permettant de savoir pourquoi un nom de fichier

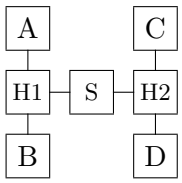
7. Du moins en première approximation...

8. Pour semer le doute, un examinateur facétieux peut ajouter que la création d'un fichier nommé `*` peut être simplifiée en se plaçant dans un répertoire vide...

commençant par `-` sera généralement considéré comme une option, alors que les caractères `>`, `*` ou `;` présents dans les noms de fichiers ne seront pas interprétés [9] dans l'exemple de la question.

Pratiquement aucun candidat ne peut avoir ce niveau d'analyse le jour de l'entretien – il faut des notions relativement avancées en administration et en programmation. La discussion met donc en évidence l'intérêt d'une bonne culture générale, et la capacité de structurer l'analyse pour rechercher les informations nécessaires à une réponse consolidée à une question loin d'être triviale.

Un peu de réseaux



Q Soit le réseau IP ci-dessus, avec A, B, C et D des ordinateurs, $H1$ et $H2$ des *hubs*, et S un *switch*.

- ☐ Un paquet émis par A à destination de $@_D$ est reçu par D
- ☐ Un paquet émis par C à destination de $@_D$ est reçu par B
- ☐ Un paquet émis par C à destination de $@_B$ est reçu par A
- ☐ A peut émettre un paquet avec $@_B$ en adresse source et $@_D$ en adresse destination, paquet qui sera reçu et traité par D

Ici, l'idée de la question est encore de distinguer le fonctionnel des aspects sécurité. Tous les paquets IP circulant sur un bus sont reçus par toutes les machines, quelle que soit l'adresse IP de destination. Mais le mécanisme coopératif consiste pour une machine à ignorer ce qui ne lui est pas destiné.

Comme indiqué en introduction, il est toujours intéressant de rappeler à un candidat que pour router un paquet, l'adresse source ne sert généralement à rien, donc elle n'est pas vérifiée par défaut par les équipements. Il est alors possible de la modifier, avec des conséquences diverses et variées (se faire passer pour quelqu'un d'autre, faire du déni de service avec amplification, etc.). C'est quand on se pose ces questions de sécurité qu'on en arrive à faire des vérifications a priori inutiles.

Un peu de C

Q Le langage C est un langage...

- ☐ machine
- ☐ de script
- ☐ compilable
- ☐ exécutable par une machine virtuelle

Q On considère un exécutable et son fichier source associé en C qui contient notamment une directive `#define`. Celle-ci est traitée :

- ☐ avant le lancement de l'exécutable
- ☐ au lancement de l'exécutable
- ☐ pendant l'exécution de l'exécutable

9. Pas par le *shell* en tout cas.

Avec l'utilisation commune des environnements de développement intégré, il est de moins en moins évident pour un développeur de faire la différence entre langage interprété et un langage compilé. Cependant, il n'est pas inutile de comprendre comment les étapes de compilation, de stockage et d'exécution d'un programme fonctionnent pour en analyser l'impact du point de vue de la sécurité.

Par exemple, on peut se demander à quoi servent les protections NX sur les pages mémoire ou le retrait du droit d'exécution sur une classe Java...

Questions ouvertes en entretien

Q Avez-vous déjà constaté/eu à traiter un incident de sécurité ?

Il est toujours intéressant, pour un candidat ayant déjà eu à gérer un système d'information, de demander s'il a déjà été confronté à un incident. Et face à une réponse négative (la plus fréquente), de provoquer une introspection sur l'existence ou non de moyens permettant de détecter une telle attaque.

Évidemment, ces questions peuvent donner lieu à une discussion intéressante, dans laquelle il pourra être question d'anti-virus et de ports USB bouchés à la colle.

Q Quelles questions poseriez-vous à un fabricant de clés USB chiffrées (avec AES-256 !) cherchant à vous vendre son produit ?

C'est le genre de situations auxquelles de nombreux élèves de la formation seront confrontées en pratique dans leurs postes. Nous n'attendons pas ici une remise en cause de la sécurité de l'algorithme AES, mais plutôt des interrogations sur le mode de chiffrement utilisé, sur le cycle de vie des clés, ou encore sur la manière dont la clé USB est déverrouillée par l'utilisateur.

3 Conclusion

La mise en place de cet entretien (et de ce questionnaire) pour la formation ESSI nous a permis d'évaluer de manière plus précise les aptitudes des candidats. En pratique, cette évaluation est souvent représentative des résultats obtenus ensuite durant la formation. L'objectif est également de préparer les futurs élèves aux attendus de la formation : une vision technique et large de la sécurité, un état d'esprit de questionnement permanent, et une rigueur dans le discours.

Enfin, il est intéressant de constater que certaines personnes arrivant et repartant très sûres d'elles, obtiennent un avis défavorable pour l'inscription à la formation, alors que d'autres, profondément déstabilisées par le questionnaire et l'entretien et persuadées d'avoir échoué, sont retenues. En pratique, l'immense majorité des candidats fait des erreurs dans le questionnaire et lors de l'entretien. Mais les candidats les plus secoués ou déçus sont ceux qui se sont rendu compte de leurs erreurs, et qui se remettent en question.