TLS-SEC

Exercices d'examen

Décembre 2015

Détection d'intrusion

Exercices proposés avec corrigé

Exercice 1 *(1 point)*

Expliquez à quoi correspond un « vrai négatif » dans le domaine de la détection d'intrusion.

Corrigé

Un vrai négatif est un cas où le système de détection d'intrusion n'émet pas d'alertes et où ceci correspond effectivement à une situation où aucune activité malveillante n'est présente dans le système d'information.

Exercice 2 *(5 points)*

Voici une signature de détection réseau utilisable avec le logiciel Snort pour détecter des messages électroniques présentant des caractéristiques spécifiques :

```
alert tcp $EXTERNAL_NET any -> $SMTP_SERVERS 25 (msg:"cyber-drrck"; flow:to_server,
established; content:"terroriste"; nocase; content:"islam"; within:30; nocase;
content:"voiture"; within:30; nocase; content:"nitrate"; within:30; nocase;
reference:nsa,765432;  classtype:string-detect; rev:2;)
```

Question 1 *(1 point)* : Expliquez sur quels critères Snort détecte un message électronique avec cette signature (type de flux réseau, caractéristiques des données) ?

On considère que la France compte 65 millions d'habitants environ. Faisons l'hypothèse que :

- chaque citoyen français utilise la messagerie électronique 5 fois par an en moyenne ;
- et parmi eux, il y a 10 individus malveillants et dangereux.

Question 2 *(1 point)* : Si la fiabilité de détection d'une signature Snort du type ci-dessus, utilisée pour repérer des communications malveillantes, est de 99,9 % : combien d'alarmes vont être générées ? Est-ce que tous les terroristes seront détectés ? Combien d'alarmes seront des faux positifs ?

Question 3 *(1 point)* : Quel taux de détection doit-on avoir pour ne plus avoir *que* de vraies alarmes ?

Question 4 *(1 point)* : Supposons que la détection soit effectuée par des mots-clefs. Supposons que les services de renseignement chargés de la mise en œuvre estiment que chaque mot-clef supplémentaire (judicieux bien évidemment) augmente la sélection des « bons » candidats d'un facteur 20. *Combien faut-il de mots-clefs* dans une signature pour atteindre le taux de détection obtenu ci-dessus (au 3 si possible, au 2 sinon). Pour mémoire : $X^n = \mathrm{e}^{n \cdot \ln(X)}$

Question 5 *(1 point)* : Discutez votre résultat.

Corrigé

1. *Les critères de détection sont : d'abord d'un point de vue réseau, il s'agit d'une connexion TCP à destination du port (serveur) n°25 (port standard des serveurs de messagerie SMTP). Ensuite, la sonde recherche dans la transmission entre le serveur et son client un flux contenant les chaînes de caractère « terroriste », « islam », « voiture » et « nitrate ». Ces recherches sont assorties de quelques options modificatrices : pas de prise en compte de la casse et distance maximale entre les mots de 30 caractères.*

2. *Compte tenu du taux indiqué, le nombre d'alarmes générées sera de*
   $65 \cdot 10^6 \times 5 \times (1 - 99{,}9\,\%) = 325\,000$
   Parmi elles, on aura 49 ou 50 vrais positifs correspondant aux messages échangés par les 10 individus effectivement malveillants. Puisqu'on fait l'hypothèse d'un tel taux de détection, on arrivera bien à identifier tous les individus et probablement tous les cinquantes messages qu'ils s'échangent (à un près). Cela fera donc un total de 324 950 faux positifs. (La très grande majorité des alarmes seront donc des fausses alarmes.)

3. *On ne veut avoir que 50 alarmes en tout. Le taux cherché est donc tel que*
   $65 \cdot 10^6 \times 5 \times (1 - \tau) = 50$ *soit* $\tau = 1 - \frac{1}{6{,}5 \cdot 10^6} \approx 99{,}9999846\,\%$

4. *Si on suppose que chaque nouveau mot-clef réduit la probabilité d'échec de la mauvaise détection d'un message par la sonde d'un facteur 20, avec n mots-clefs, le taux de détection est de* $1 - \frac{1}{20^n}$ *et donc pour la valeur identifiée précédemment*
   $\tau = 1 - \frac{1}{6{,}5 \cdot 10^6} = 1 - \frac{1}{20^n}$ *d'où* $20^n = 6{,}5 \cdot 10^6$ *et donc*[1]
   $n = \frac{\ln(6{,}5 \cdot 10^6)}{\ln(20)} \approx \frac{15{,}68}{2{,}99} \approx 5{,}24$ *. Ceci nous amène à la conclusion que 6 de ces mots-clefs suffisent pour obtenir le taux de sélection nécessaire à une détection « parfaite ».*

5. *Une signature Snort recherchant 6 mots-clefs dans les échanges de messagerie électronique qui peut permettre d'identifier les individus malveillants avec une fiabilité quasi absolue ? A l'approche de Noël, nous nous surprenons à rêver. Néanmoins ici le résultat est surtout extrêmement douteux. En fait, il est certain que l'usage des différents mots dans un message est corrélé (ne serait-ce que par le sens du message dans son ensemble) et qu'il donc impossible qu'il soit chacun un nouveau facteur de discrimination indépendant (qui plus est suffisamment efficace pour*

[1] Comme $20^n = e^{n \cdot \ln(20)}$ et en prenant le logarithme.

distinguer à coup sûr un élément parmi 20). Ensuite, il est fort probable que les citoyens malveillants sont précisément ceux qui évitent d'utiliser des mots-clefs repérables (sans même parler de chiffrement, la simple utilisation d'un mot pour un autre suffirait à tromper la sonde réseau) ; on doit donc aussi s'attendre à des faux négatifs. En fait, globalement, à part le nombre d'habitants, toutes les hypothèses faites à la question 4 sont extrêmement douteuses.
Comme de plus elles conduisent à installer des systèmes de surveillance réseau systématique et à les paramétrer par des listes de mots-clefs (comment et par qui seront-ils choisis ?), sans trop se préoccuper de faire sur le terrain de distinction entre vraies et fausses alarmes ; on peut aussi demander si ceux qui formuleraient ce genre d'hypothèses ne sont pas dangereux également. En tout cas, ils ne semblent pas contribuer de manière crédible à résoudre le problème d'identification d'actions réellement malveillantes.