Identification du module

Numéro de module	683	
Titre	Analyser et interpréter des ensembles de données	
Compétence	Inspecter des données brutes et des ensembles de données quant à la présence d'informations critiques et déterminantes pour la sécurité, plausibiliser les résultats et les exploiter de façon probante en adéquation avec le public cible.	
Objectifs opérationnels	1	Collecter des données brutes non structurées, semi-structurées ou structurées qui sont pertinentes pour la sécurité à partir des systèmes, des applications, des solutions de surveillance et de protection d'une organisation.
	2	Masquer, pseudonymiser ou anonymiser des données brutes sensibles si nécessaire et conformément à la protection des données.
	3	Programmer des scripts et des outils pour évaluer, traiter et représenter de grands ensembles de données.
	4	Interroger des bases de données et mettre en forme les résultats obtenus.
	5	Inspecter les données quant à la présence d'anomalies, d'indicateurs ou de motifs spécifiques à des incidents de sécurité potentiels.
	6	Plausibiliser les données obtenues, identifier et filtrer les faux positifs et étoffer le contenu informatif des résultats en enrichissant les données brutes avec des informations complémentaires lisibles.
	7	Evaluer l'analyse des données, consigner les résultats dans un rapport final probant et présenter les résultats aux décideurs en adéquation avec le groupe cible.
Domaine de compétence	Security/Risk Management	
Objet	Données brutes et ensembles de données relevant de la sécurité et issus de systèmes, d'applications, de solutions de surveillance et de protection d'une organisation.	
Version du module	1.0	
Créé le	11.02.2021	

Connaissances opérationnelles nécessaires

Les connaissances opérationnelles nécessaires décrivent les savoirs qui contribuent à l'exécution compétente des actions d'un module. La définition de ces connaissances est fournie à titre indicatif et n'est pas exhaustive. La concrétisation des objectifs de formation et la définition du chemin d'apprentissage en vue d'acquérir ces compétences incombent aux prestataires de formation.

Numéro de module	683
Titre	Analyser et interpréter des ensembles de données
Compétence	Inspecter des données brutes et des ensembles de données quant à la présence d'informations critiques et déterminantes pour la sécurité, plausibiliser les résultats et les exploiter de façon probante en adéquation avec le public cible.

Objectifs opérationnels et connaissances opérationnelles nécessaires

1	1.1	Connaître des exemples typiques de données non structurées, semi-structurées ou structurées et pouvoir expliquer leurs différences et leurs spécificités au regard de l'analyse des données.
	1.2	Connaître des sources usuelles de données pertinentes pour la sécurité dans l'infrastructure informatique d'une organisation (p. ex. SIEM, fichiers log et fichiers journaux d'appareils et d'applications, e-mail, messages instantanés, bases de données d'applications).
	1.3	Connaître des formats de données texte et binaires pour l'échange et la sérialisation des données (p. ex. CSV, XML, JSON, BSON, YAML, HDF) et pouvoir expliquer leurs spécificités, leurs différences et leur pertinence pour les analyses de données.
	1.4	Connaître divers standards et formats de fichiers journaux (p. ex. Syslog, journal des événements Windows, fichiers log de serveurs Web).
	1.5	Connaître des outils appropriés pour la gestion des logs (p. ex. Splunk, Papertrail, Loggly, GrayLog, Logstash) et pouvoir expliquer leurs fonctionnalités (p. ex. collecte, indexation, analyse, visualisation).
2	2.1	Connaître l'importance du masquage, de la pseudonymisation et de l'anonymisation statiques et dynamiques des données et pouvoir expliquer leurs différences ainsi que les défis à relever lors de leur mise en œuvre.
	2.2	Connaître diverses méthodes de masquage, de pseudonymisation et d'anonymisation des données (p. ex. suppression, substitution, hachage, réarrangement de données [shuffling], méthode de variance, méthodes de chiffrement) et pouvoir expliquer leur principe de fonctionnement.
	2.3	Connaître diverses procédures d'anonymisation (p. ex. k-anonymité, l-diversité, confidentialité différentielle, Diffix) et pouvoir expliquer leur niveau de sécurisation face à la désanonymisation ou réidentification et leur influence sur la qualité et la valeur informative des données brutes.
	2.4	Connaître les dispositions légales de la protection des données en matière de pseudonymisation et d'anonymisation des données à caractère personnel et pouvoir expliquer comment respecter les principes de licéité, de proportionnalité, de finalité et de transparence à l'aide d'exemples d'application typiques.

Connaissances opérationnelles nécessaires

	2.5	Connaître des outils appropriés de masquage, de pseudonymisation et d'anonymisation d'ensembles de données (p. ex. Amnesia, Anonimatron, ARX).
3	3.1	Connaître des expressions régulières et pouvoir expliquer leur pertinence dans les analyses de données.
	3.2	Connaître des commandes et des concepts pertinents (p. ex. entrées et sorties, filtres, tubes [pipes], redirection) des interpréteurs de commande des systèmes d'exploitation Windows (cmd, PowerShell) et Unix/Linux (p. ex. bash, ksh, zsh) et pouvoir expliquer leurs possibilités et leurs limites dans le contexte de l'analyse des données.
	3.3	Connaître la syntaxe du langage de programmation Python, les bibliothèques de référence (p. ex. NumPy, SciPy, Pandas, Matplotlib) et les outils usuels (p. ex. IPython, Jupyter Notebooks) pour l'analyse et la représentation des données.
	3.4	Connaître des langages de programmation alternatifs à Python (p. ex. R, Scala, Julia, MATLAB) et pouvoir expliquer leur pertinence dans le contexte de l'analyse des données.
4	4.1	Connaître le concept de base de données relationnelle et SQL, pouvoir citer des exemples d'application et des technologies typiques (p. ex. MySQL, serveur SQL, PostgreSQL) et expliquer leurs avantages et leurs inconvénients dans le traitement de grandes quantités de données.
	4.2	Connaître des technologies courantes de bases de données non relationnelles NoSQL (p. ex. Cassandra, BigTable, MongoDB, CouchDB, Riak) et pouvoir expliquer leur modèle de données (orienté colonne, orienté document, orienté graphe, valeur-clé), les possibilités d'effectuer des requêtes (p. ex. UnQL, méthodes objet) ainsi que leurs avantages et leurs inconvénients pour le traitement de grandes quantités de données.
5	5.1	Connaître des techniques et des concepts pertinents pour l'analyse de grandes quantités de données (p. ex. reconnaissance de motifs, A/B testing, analyse de séries temporelles, corrélation statistique et régression, apprentissage automatique [machine learning].
	5.2	Connaître des indicateurs typiques de compromission (IoC), p. ex. valeurs de hachage, signatures, noms de domaine, adresses IP, URL, adresses e-mail, X-Mailer, HTTP User Agent).
	5.3	Connaître des anomalies typiques relatives à la sécurité (p. ex. écarts dans l'utilisation de la bande passante, anomalies au niveau des protocoles/ports).
6	6.1	Connaître l'importance des faux positifs pour l'évaluation des ensembles de données et pouvoir expliquer des exemples typiques.
	6.2	Connaître des sources usuelles d'informations complémentaires en vue d'enrichir les données (p. ex. DHCP, AD, inventaire, base de données de configuration).
	6.3	Connaître des commandes pertinentes pour recueillir des informations complémentaires (p. ex. nslookup, whois, trace/tracert) et pouvoir expliquer le contenu de ces informations.
7	7.1	Connaître des méthodes et des techniques appropriées de synthèse et de présentation des informations (p. ex. réduction des données, création de ratios, tableaux de fréquence et histogrammes, agrégation au moyen de tableaux croisés et de tableaux croisés dynamiques, diagramme de corrélation, analyse de séries temporelles et analyse des tendances).

Connaissances opérationnelles nécessaires

	7.2	Connaître les principaux contenus d'un rapport final d'analyse (p. ex. synthèse, contexte, objet de l'analyse, méthodes, résultats, constats, options d'action, recommandation avec justification) et pouvoir expliquer leur contribution à la prise de décisions.
	7.3	Connaître les contenus et la structure d'une bonne présentation et pouvoir expliquer en quoi ses propres compétences en termes d'expression et de comportement influencent le travail de persuasion.

Version du module	1.0
Créé le	11.02.2021