Veille technologique : l’IA et les risques sur la Cybersécurité

Sujet : L'Intelligence Artificielle au service de la cybercriminalité

Cadre : Épreuve de BTS SIO - Option SISR

Table des matières

I. L’AUTOMATISATION DE L’INGÉNIERIE SOCIALE

L'IA permet aux attaquants de mener des campagnes de phishing à une échelle industrielle avec une précision inédite.

1.1 Le Phishing Génératif et "WormGPT"

Les attaquants utilisent des modèles de langage (LLM) non censurés pour rédiger des courriels de phishing sans aucune faute, imitant parfaitement le ton de grandes institutions ou de collaborateurs internes.

- **Le Risque :** Les indices classiques (fautes d'orthographe, syntaxe) disparaissent. Chaque message est unique, ce qui rend la détection par les filtres antispam classiques très difficile.
- **Lien Source :** SlashNext - L'essor de WormGPT et des outils d'IA malveillants

1.2 L'usurpation d'identité par Deepfake (Vishing & Vidéo)

Le clonage vocal et vidéo permet de simuler la présence de dirigeants lors de procédures de validation critique.

- **Exemple concret :** Une entreprise a perdu 25 millions de dollars après qu'un employé a participé à une visioconférence où tous ses interlocuteurs étaient des clones numériques.
- **Lien Source :** CNN - Fraude massive par Deepfake (25M$)
-

II. L’ÉVOLUTION DES MALWARES ET DE L'EXPLOITATION

L'IA modifie la manière dont les codes malveillants contournent les dispositifs de sécurité réseau.

2.1 Les Malwares Polymorphes (Projet BlackMamba)

Contrairement aux virus classiques, un malware utilisant l'IA peut réécrire son propre code à chaque exécution pour modifier sa signature numérique (hash).

- **Impact SISR :** Les antivirus basés sur des signatures sont obsolètes. Le passage à des solutions **EDR (Endpoint Detection and Response)** analysant le comportement est désormais une nécessité technique.
- **Lien Source :** CyberArk - BlackMamba : Générer des malwares polymorphes avec l'IA

2.2 Le cassage de mots de passe par Deep Learning (PassGAN)

L'outil **PassGAN** utilise des réseaux de neurones pour analyser des milliards de combinaisons issues de fuites de données passées afin de prédire les nouveaux mots de passe.

- **Statistique :** L'IA peut casser environ 51% des mots de passe courants en moins d'une minute.
- **Lien Source :** HomeSecurityHeroes - Comment l'IA craque vos mots de passe

III. VULNÉRABILITÉS PROPRES AUX INFRASTRUCTURES IA

Le déploiement d'outils d'IA au sein du SI crée une nouvelle surface d'attaque pour l'administrateur.

3.1 Le "Prompt Injection" (Injection d'instructions)

C'est l'équivalent moderne de l'injection SQL appliqué aux modèles de langage. Un attaquant "manipule" l'IA pour lui faire ignorer ses consignes de sécurité et extraire des données sensibles.

- **Référentiel :** Le projet OWASP a dédié un classement spécifique pour ces nouvelles failles logicielles.
- **Lien Source :** OWASP Top 10 pour les applications LLM

3.2 L'empoisonnement des données (Data Poisoning)

Consiste à injecter des données erronées dans les flux réseaux pour que les outils de détection (SIEM) ou les modèles de défense apprennent à ignorer des activités malveillantes réelles.

IV. SYNTHÈSE DES SOLUTIONS DE DÉFENSE (RÔLE SISR)

Type de Menace	**Solution d'Infrastructure Recommandée**
Phishing / Vishing	Authentification Forte (**MFA / Clés FIDO2**) et protocoles de validation.
Malware Polymorphe	Analyse comportementale via **EDR / XDR**.
Crack Pass IA	Adoption des **Passkeys** et abandon des mots de passe statiques.
Injections IA	Mise en place de passerelles sécurisées et filtrage des entrées/sorties LLM.

Sources de Veille Stratégique :

1. **ANSSI** : ssi.gouv.fr (Référence française)
2. **The Hacker News** : thehackernews.com (Actualité technique mondiale)
3. **CrowdStrike** : crowdstrike.com (Intelligence sur les menaces)