Comment détecter et lutter contre les comportements procéduraux de bots malveillant sur Internet ?

William Piron
M2 MIAGE Classique

Présentation du sujet de mémoire
avant validation définitive

Tuteur : M. Fabrice Legond-Aubry, Maître de conférence

Octobre 2017

Contents

Présentation et raison du sujet

Avec la démocratisation progressive depuis le début des années 90 de l'utilisation de bots informatiques sur Internet, de nombreuses problématiques se posent. Au début majoritairement utilisés comme chatbots, c'est à dire pour animer des chats en lignes et y fournir divers services, les bots ont été récupérés et utilisés à de nombreuses fins depuis.

De nombreux problèmes se posent actuellement vis à vis des bots dans des contextes comme les réseaux sociaux, les sites d'informations et bien d'autres. Ils sont utilisés pour effectuer des tâches répétitives dans un but malveillant, notamment diffuser de fausses informations, envoyer des mails de spam, effectuer des attaques par déni de service ou effectuer des actions de répétitives contre rétribution réelle sur des jeux en ligne.

Particulièrement vulnérables aux bots, on trouve les réseaux sociaux et les objets connectés. Dans le premier cas, il s'agit de spambots, parfois de particuliers, notamment pour effectuer des arnaques ou du phishing, et de plus en plus gouvernementaux ou mis en place par des organismes politiques, afin d'orienter les opinions ou diffuser de fausses informations. Dans le second, ces bost cherchent des objets connectés non sécurisé ou très peu, et les détourne afin de servir de relais à des attaques de dénis de service.

Afin de lutter contre l'utilisation détournée de cet outil, de nombreuses solutions ont été mises en place afin de détecter et d'agir contre ces comportements répétitifs et identifiables. Celles-ci se reposent le plus souvent sur la détection de ces actions, en repérant des schémas d'exécution, afin de cibler les utilisateurs suspectés d'être des bots, ou tracent les données de ces cibles pour identifier leur source et déterminer s'il s'agit de personnes réelles ou non.

Ce mémoire aura pour but de réfléchir aux solutions existantes, leur avantages et limitations, de tester certaines d'entre elles en condition réelle, et de proposer des pistes complémentaires de réflexions pour améliorer ou mettre au point un complément de solution de lutte contre ces bots.

Utilité de la démarche

Devant le nombre croissant de soucis liés aux fausses informations relayées sur les réseaux sociaux, aux attaques DDOS réalisées à l'aide de bots ou de vagues de spam, de nombreux acteurs cherche à mettre en place une solution pour filtrer sur leurs services les données provenant de bots.

Une grande partie du travail dans la lutte contre les bots malveillant consiste à déterminer quel type de comportement est suspicieux, et comment le bloquer. Les réseaux sociaux sont

remplis de comptes récents ne parlant que d'un sujet, ou ne participant qu'à certains types de discussions. D'autres sont mis en place en avance, avec une activité minime jusqu'à leur entrée en activitée concrète, que ce soit pour effectuer un déni de service ou lancer une campagne de spam ou phishing.

De nombreuses solutions sont mises en place sans pouvoir complètement bloquer leur action, ou sont rapidement contournées. D'autres ont un impact trop grand sur les utilisateurs normaux des services et entraînent la chute de ces derniers.

Il devient nécessaire de réfléchir à la mise en place de système pour sécuriser aussi bien les sites internet, les jeux en ligne, mais surtout les objets connectés, actuellement très faibles en matière de sécurité. Le gouvernement estime d'ailleurs qu'il s'agit d'un des secteurs critiques et porteurs à l'horizon 2020, montrant que ce thème précis, pouvant être étendu, est primordial dans un futur proche.

Réflexions et propositions d'expérimentations

Ma proposition par le biais de ce mémoire sera de déterminer, à mon échelle, un environnement propice à des tests de technologies existantes (par exemple, héberger un ou plusieurs sites web, mettre en place certaines mesures anti-bots et les tester) et de les comparer en conditions réelles. Le but final sera soit la détermination de la meilleure solution selon le contexte et l'échelle, soit la proposition d'une nouvelle méthode ou de méthodes conjointes. Une réflexion sur les limites de ces technologies ainsi que sur leur impact sur les utilisateurs sera également menée.

Le but de ce mémoire ne sera pas de proposer un nouvel algorithme révolutionnaire, ou une solution parfaite, mais de réfléchir sur les solutions actuelles, sur leurs implémentations, sur leur limitations et leur avenir, en proposant des pistes et des résultats d'expérimentations.

Mise en place de tests

Afin d'appuyer cette réflexion, il sera mis en place différentes méthodes pour tester les solutions existantes dans la lutte contre les bots malveillants.

Une recherche approfondie des techniques existantes et de leur limite sera effectuée. Elle sera suivi de la récupération dans le but de tests d'algorithmes open-source ou gratuits, puis de leur déploiement sur différentes sites web qui auront été codés et déployés en avance.

Une première procédure de tests sera la suivante. Des bots tenteront d'accéder à ces sites dans le but d'effectuer des dénis de service, à plusieurs reprises et dans des nombres différents. Le but sera de refuser les connections ou de les limiter à la suite de la détection de leur comportement, afin de vérifier que les algorithmes testés sont efficaces. Il faudra également tester l'efficacité des détecteurs de spambots, notamment sur les réseaux sociaux. Afin d'éviter tout problème légal, ils ne seront pas déployés sur de vrais réseaux sociaux, il faudra donc mettre en place un environnement en monde fermé afin de les simuler. Enfin, dans le but de collecter des informations pratiques sur la mise en place de ce type d'algorithmes, il peut être utile de contacter différents acteurs français du jeu vidéo multijoueurs en ligne (notamment la société Ankama) concernant leur lutte contre les bots déployés par des tiers sur leurs serveurs de jeu, leur impact, et l'efficacité de leurs méthodes.

D'autres procédures de tests seront mises en place en fonction des résultats de la première, de mes recherches, et de mes échanges avec les enseignants de la MIAGE. On peut notamment penser à tester les attaques contre les objets connectés, ou d'autres variantes moins connues découvertes pendant la rédaction du mémoire. Selon les retours de mon tuteur, il est également possible de centrer le sujet du mémoire sur un point plus précis, ou de revoir complètement les procédures de tests.