

Systeme d'analyse des spams à des fins de renseignement criminel : traitement des e-mails

Problématique

Depuis un certain temps, les escroqueries sur Internet prennent une plus grande ampleur années après années. Les escrocs inventent constamment d'autres histoires plus surnoisées les unes que les autres afin de soutirer de l'argent illégalement au grand nombre de victimes potentielles qu'offre Internet.

La police, disposant actuellement de peu de connaissances sur ce sujet, a décidé de mettre en place un système permettant de réunir un maximum d'informations à leur sujet. Ce travail étant actuellement effectué manuellement, il demande énormément de temps pour récupérer un minimum d'informations.

Le système final devra pouvoir traiter toutes les sortes d'escroquerie, les plus courantes étant la *Nigerian Connection*, le *Phishing*, la *Loterie* et les *Appels de dons*.

La solution

Ce projet consiste donc à créer un système qui, de manière autonome, prélève des e-mails sur certains comptes de messagerie, récupère un maximum d'informations dans ces messages, informations qui sont stockées et destinées à une analyse ultérieure. Chaque e-mail est finalement classifié dans un certain type d'escroquerie. Le système est décomposé en cinq phases expliquées ci-dessous.

La première d'entre elles, appelée « Récupération des e-mails », consiste à se connecter sur divers comptes de messagerie et d'y récupérer les e-mails stockés.

Le but de la seconde phase, « Nettoyage de l'information », est de récupérer les e-mails intégrés dans le corps d'un e-mail parent. Cette phase utilise pour cette opération plusieurs algorithmes différents.

La troisième phase, « Extraction des informations », consiste à l'extraction d'informations depuis les e-mails récupérés sur les divers comptes de messagerie ou extraits d'autres messages.

La quatrième, « Classification des données », classe automatiquement les e-mails dans une catégorie d'escroquerie.

La dernière phase, « Analyse métier », propose les outils nécessaires aux organes de police ainsi qu'aux analystes pour effectuer diverses sortes d'analyses sur les données du système.

Dans ce projet, seules les trois premières phases ont été développées. Les deux phases restantes feront l'objet d'un travail ultérieur.

Au final, l'analyse des informations ainsi extraites permettront à la police scientifique de mieux connaître les agissements des auteurs de ces escroqueries et, dans l'idéal, de remonter à la source de celles-ci.

