DOCTORATS INDUSTRIALS

EL PLA DE DOCTORATS INDUSTRIALS

PROJECTE DE DOCTORAT INDUSTRIAL EXPEDIENT 2016 DI 091

DADES DE L'EMPRESA I DE L'ENTORN ACADÈMIC

Títol del projecte
Disseny d'un sistema de classificació de pàgines web i programari maliciós
Empresa
Blueliv, Leap in Value
Responsable de l'empresa
Gerard Cervello Garcia
Universitat
Universitat de Lleida
Director/a de tesi
Jordi Planes Cid
Treballador/a de l'empresa i doctorand/a
Daniel Gibert Llaurado

BREU DESCRIPCIÓ DEL PROJECTE DE RECERCA

A causa del ràpid avenç en les tècniques de descobriment d'informació, l'aprenentatge automàtic i la mineria de dades continua jugant un rol major en l'àmbit de la ciberseguretat. Tanmateix, el nombre de pàgines web en Internet continua incrementant significativament any rere any. En conseqüència, l'anàlisi i classificació de pàgines web a partir del seu contingut ha esdevingut una tasca molt important per a poder protegir els usuaris d'accedir a pàgines web amb continguts maliciosos. Una pàgina web maliciosa es pot descriure com un lloc web que intenta dur a terme qualsevol tipus d'acció no desitjada com per exemple, instal·lar software maliciós en el dispositiu o robar informació sensible sobre l'usuari.

L'objectiu fonamental d'aquest projecte és el disseny i implementació d'un sistema de classificació de pàgines web a partir del seu contingut i característiques. Els objectius del projecte són:
-Dissenyar un sistema capaç de detectar i classificar pàgines web malicioses.
-Dissenyar un sistema escalable que sigui capaç de processar grans quantitats de dades en temps real i aplicar tècniques de mineria de dades per a analitzar les pàgines web.
-Estudiar com es poden aplicar tècniques d'aprenentatge automàtic, especialment tècniques de deep learning, per a resoldre la tasca actual.

El sistema haurà de ser capaç de classificar les pàgines web en almenys les següents classes:
-C&Cs botnet servers. Una xarxa de bots està composta per varis dispositius connectats mitjançant Internet capaços de comunicar-se entre ells i amb altres dispositius per a coordinar accions mitjançant comandaments. Els llocs web pertanyent a aquesta classe seran aquells responsables d'executar els comandaments i emmagatzemar la informació rebuda de la botnet (xarxa de bots).
-Phishing. Els llocs web classificats com a phishing són aquells que intenten obtenir informació sensible sobre l'usuari com per exemple, noms d'usuari, contrasenyes o informació de targetes de crèdit.
-Malware. Els llocs web classificats com a malware seran aquells responsables de descarregar e instal·lar software maliciós.
-Exploit Kits. Un exploit kit és un servidor web en el qual s'ha instal•lat software per a identificar i explotar (aprofitar) vulnerabilitats dels dispositius que interactuïn amb ell.

A més a més, si el lloc web és utilitzat per a la descàrrega de software maliciós, el sistema haurà de ser capaç d'analitzar i agrupar aquest software en famílies a partir del seu contingut i comportament.