Zwalczanie spamu zyskało wymiar… lingwistyczny

2 grudnia 2017

Nawet najsprytniejsze boty, rozsyłające spam i wirusy, mogą być wykryte. Zdradza je dialekt, za pomocą którego komunikują się z serwerem pocztowym. Badacze w NASK PIB pracują nad bazą danych takich dialektów.

„Poszukujemy zdalnych pracowników…", „odbierz rabat…", „Usprawnij swoje życie seksualne…" – większość z nas ma już serdecznie dosyć spamu. Niechcianej poczty, która zapycha nam skrzynki mailowe i zwiększa ryzyko, że przegapimy jakąś ważną wiadomość, na którą czekamy. Spam bywa nie tylko irytujący, ale czasem jest też niebezpieczny, np. kiedy klikniemy w jakiś niegodny zaufania link i na naszym komputerze zainstaluje się bez naszej wiedzy groźne oprogramowanie.

Dlatego specjaliści poszukują kolejnych sposobów identyfikowania seryjnie rozsyłanej korespondencji. Jednym z takich sposobów jest analiza dialektu, czyli drobnych różnic w użyciu protokołu komunikacyjnego SMTP. O badaniach dotyczących tego tematu poinformowali przedstawiciele NASK Państwowego Instytutu Badawczego w przesłanym PAP komunikacie.

"Protokół SMTP, używany przez programy pocztowe do komunikacji z serwerem, dopuszcza pewną swobodę kodowania. Skutek jest taki, że poszczególne klienty poczty różnią się od siebie, a różnice w ich komunikacji z serwerem można zaobserwować i wykorzystać do identyfikowania źródła korespondencji" – wyjaśnia Piotr Bazydło z Zespołu Metod Bezpieczeństwa Sieci w NASK PIB.

Protokół SMTP pozwala autorowi programu wysyłającemu pocztę na pewną dowolność przy pisaniu kodu. Dodatkowe spacje, wielkie lub małe litery, przytoczenie nazwy domeny zamiast adresu IP nadawcy i inne, nie stwarzają protokołowi problemów. O ile kod

trzyma się ogólnych ram, zostanie wykonany i komunikacja przebiegnie prawidłowo. Jednak analizując tę komunikację – za pomocą programu, zainstalowanego na serwerze, można te różnice zobaczyć i potraktować jak odciski palców.

Dialektem różnią się między sobą używane powszechnie programy pocztowe. Podglądając komunikację klienta z serwerem można ustalić czy e-mail został wysłany z programu Outlook, Thunderbird, serwera Gmail czy innego. Gdy analiza wykaże, że kod nie odpowiada żadnemu ze znanych programów lub stanowi kombinację więcej niż jednego z nich, to widać, że pocztę wysłano z użyciem mniej oficjalnego narzędzia, być może bota.

“Jako pierwsi zaobserwowali różnice w dialektach autorzy pracy opublikowanej w USA w 2012 r. My postanowiliśmy kontynuować badania nad wykorzystaniem różnic w dialektach do analizy spamu i botnetów, które go rozsyłają. W naszym wydaniu to narzędzie w dużej mierze kryminalistyczne, które może dostarczać dane na potrzeby policji i wymiaru sprawiedliwości” – zaznacza Piotr Bazydło.

Narzędzia tego można też jednak użyć do stworzenia “czarnej listy” adresów, z których wysyłana jest groźna korespondencja, co może pomóc uaktualniać filtry antyspamowe.

Wykorzystując różnice w wymianie komend między serwerem a klientem oraz inne dane, przychodzące wraz z wiadomością, badacze potrafią nie tylko wykryć pocztę pochodzącą z podejrzanego źródła, ale też coraz precyzyjniej określić, jakie to źródło, np. ustalić, jaki typ botneta wysyła dany spam.

Dane do swoich badań naukowcy zbierają przy pomocy tzw. spamtrapa, czyli komputera udającego serwer pocztowy. Pomagają one przyłapać boty, rozsyłające pocztę na losowe adresy.

Takie i inne pułapki znajdą się w systemie wczesnego ostrzegania przed cyberzagrożeniami, budowanym w ramach międzynarodowego programu badawczego SISSDEN (Secure

Information Sharing Sensor Delivery event Network). NASK PIB jest liderem tego programu.

W konsorcjum zaangażowanych jest ponadto siedem innych instytucji z różnych krajów Unii Europejskiej. SISSDEN finansowany jest w ramach programu Unii Europejskiej Horyzont 2020. Docelowo za jego sprawą ma powstać sieć przynajmniej 100 serwerów (przynajmniej po jednym w każdym kraju członkowskim), służących do wykrywania i analizowania niebezpiecznych aktywności w sieci.

Źródło: NaukawPolsce.PAP.pl