Graph-based Malware Distributors Detection

Andrei Venzhega
Web-search manager
http://firefox.mozil1a.ru/

Mozilla Firefox с «Яндекс.Бар»

В новый бета-версии Mozilla Firefox, который можно будет сканировать и попробовать новые функции и возможности, которые предустановлены и доработаны программисты и простые пользователи. Ждем новых обновлений.
Malware: search traffic

Search engines are the main source of traffic on malware distributors websites.
Malware: overview

Symantec annual report 2011:
• About 403 millions new types of malicious software were detected in 2011 (41% more than 2010)

• Signature-based methods can’t deal with unique signature malware.
Bipartite graph

website 1

website 2

website 3

website 4

website 5

website 6

filehosting 1

filehosting 2

filehosting 3

filehosting 4
Graph initialization

\[ G = (V, E, W) \]

site: \( S \subset V \); filehosting: \( F \subset V \);  

\( W : w(s, f) \in [0, 1]; \ s \in S; \ f \in F; \)

\( M : m(v) = P(\text{v is bad}) \in [0, 1]; \ v \in V. \)
Malwariness rank

website rank: \[ m_k(s_i) = \sum_{j \in N_i} \frac{w_s(s_i, f_j) m_{k-1}(f_j)}{W_s(s_i)}; \]

filehosting rank: \[ m_k(f_j) = \sum_{i \in N_j} \frac{w_f(s_i, f_j) m_{k-1}(s_i)}{W_f(f_i)}; \]

sum of edges weights:

\[ W_s(s_i) = \sum_{j \in N_i} w_s(s_i, f_j); \quad W_f(f_j) = \sum_{i \in N_j} w_f(s_i, f_j); \]

\( N_i \) – set of all neighbors of vertex \( v_i \);

the stop condition on \( k \) iteration: \[ \| m_k(v) - m_{(k+1)}(v) \| < \varepsilon; \quad \forall v \in V; \]
Method visualization. Prior set. Step 1-1
Method visualization. Step 1-2
Method visualization. Step 2-1
Method visualization. Step 2-2
Results

Method detected filehostings (FH) distributing malware (MD) files with:
97% - precision
60% - recall (strongly depends from the initialization set)

2.1 times decrease of the of malware distributors in top10 search results

Only 9% of files from malware FH at the moment of their detection were in anti-virus signature bases.

More than 90% of files from these FH were in anti-virus signature bases after 2 weeks.

1 week is the minimum lifetime of FH, an average lifetime of MD websites is much longer.
Andrei Venzhega
web-search manager
a.venzhega@ya.ru