

LeakedOut: las redes sociales que te atrapan

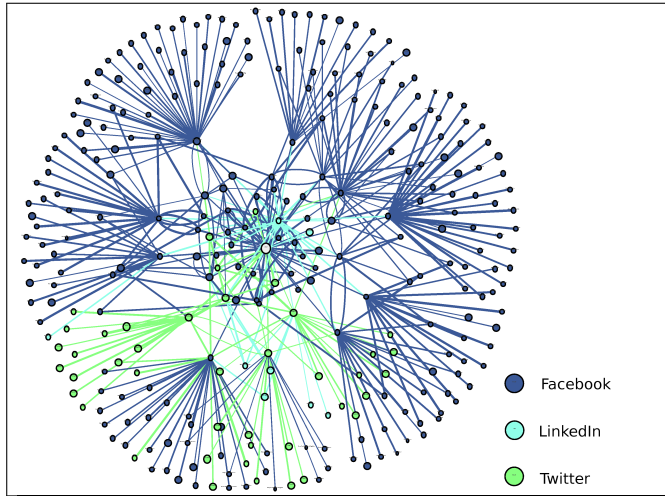
José I. Orlicki

CoreLabs - Core Security Technologies
PhD program - ITBA (BA-Con 2008 Buenos Aires)

8 de octubre de 2008



Introducción - Usted esta aquí!



Quien soy

- ▶ Mi edad es 26.
- ▶ Soy de Buenos Aires, Argentina (donde usted esta ubicado).
- ▶ Científico computacional.
- ▶ Investigador en CoreLabs por 2 temporadas. Core vende programas y servicios de pen-test.
- ▶ Estudiante de doctorado en ITBA.
- ▶ Intereses principales: Inteligencia de fuentes abiertas (OSINT), Simulacion y planeo de ataques, Análisis de redes complejas.
- ▶ Mayormente inofensivo, interesado en redes sociales luego de ser capturado por LinkedIn.



Temas de la charla

- ▶ No trata de WebApps Security, ver charla previa de BH-USA "Black Hat 2008: **Satan** Is On My Friends List", sección de OpenSocial recomendada.

Temas de la charla

- ▶ No trata de WebApps Security, ver charla previa de BH-USA "Black Hat 2008: **Satan** Is On My Friends List", sección de OpenSocial recomendada.
- ▶ No trata de cerrar tu cuenta de Facebook, igual lo puedes hacer mandando un *correo* a la gente adecuada.

Temas de la charla

- ▶ No trata de WebApps Security, ver charla previa de BH-USA "Black Hat 2008: **Satan** Is On My Friends List", sección de OpenSocial recomendada.
- ▶ No trata de cerrar tu cuenta de Facebook, igual lo puedes hacer mandando un *correo* a la gente adecuada.
- ▶ Inteligencia de fuentes abiertas (**OSINT**): procesar info **publica/filtrada** para generar inteligencia y **actuar**. (No la GPL!)

Temas de la charla

- ▶ No trata de WebApps Security, ver charla previa de BH-USA "Black Hat 2008: **Satan** Is On My Friends List", sección de OpenSocial recomendada.
- ▶ No trata de cerrar tu cuenta de Facebook, igual lo puedes hacer mandando un *correo* a la gente adecuada.
- ▶ Inteligencia de fuentes abiertas (**OSINT**): procesar info **publica/filtrada** para generar inteligencia y **actuar**. (No la GPL!)
- ▶ Centrada en Internet, Motores de búsqueda, Servicios de Redes **Sociales**, IMs, Web 0.2, etc.

Temas de la charla

- ▶ No trata de WebApps Security, ver charla previa de BH-USA "Black Hat 2008: **Satan** Is On My Friends List", sección de OpenSocial recomendada.
- ▶ No trata de cerrar tu cuenta de Facebook, igual lo puedes hacer mandando un *correo* a la gente adecuada.
- ▶ Inteligencia de fuentes abiertas (**OSINT**): procesar info **publica/filtrada** para generar inteligencia y **actuar**. (No la GPL!)
- ▶ Centrada en Internet, Motores de búsqueda, Servicios de Redes **Sociales**, IMs, Web 0.2, etc.
- ▶ Discute un modesto **prototipo** llamado **Exomind**.



Cual es el aspecto divertido?

- ▶ Navegar **gráficos** a color de puntos y líneas.

Cual es el aspecto divertido?

- ▶ Navegar **gráficos** a color de puntos y líneas.
- ▶ Aprender a controlar tu perfil **filtrado**.

Cual es el aspecto divertido?

- ▶ Navegar **gráficos** a color de puntos y líneas.
- ▶ Aprender a controlar tu perfil **filtrado**.
- ▶ Construir **replicantes** chatbots luego de ver *Bladerunner* (corte del director).

Cual es el aspecto divertido?

- ▶ Navegar **gráficos** a color de puntos y líneas.
- ▶ Aprender a controlar tu perfil **filtrado**.
- ▶ Construir **replicantes** chatbots luego de ver *Bladerunner* (corte del director).
- ▶ Tendrás permitido hablar hype luego de esta charla, usando expresiones como **Computación Nube** o **Economía de la Atención**.

Cual es el aspecto divertido?

- ▶ Navegar **gráficos** a color de puntos y líneas.
- ▶ Aprender a controlar tu perfil **filtrado**.
- ▶ Construir **replicantes** chatbots luego de ver *Bladerunner* (corte del director).
- ▶ Tendrás permitido hablar hype luego de esta charla, usando expresiones como **Computación Nube** o **Economía de la Atención**.
- ▶ Afortunadamente **comprender amenazas** reales a la privacidad personales y de tu organización.

Cual es el aspecto divertido?

- ▶ Navegar **gráficos** a color de puntos y líneas.
- ▶ Aprender a controlar tu perfil **filtrado**.
- ▶ Construir **replicantes** chatbots luego de ver *Bladerunner* (corte del director).
- ▶ Tendrás permitido hablar hype luego de esta charla, usando expresiones como **Computación Nube** o **Economía de la Atención**.
- ▶ Afortunadamente **comprender amenazas** reales a la privacidad personales y de tu organización.
- ▶ En resumen: ahora con la **Wikinomía** y la Web 0.2 existen grandes cantidades de info pública/filtrada disponibles, tal vez necesites medir las consecuencias !?!



Client-side es un subgénero de pentest

OSINT automática ya forma parte de los pentests profesionales.

- ▶ **Phishing** and explotación de apls. cliente llevaron al desarrollo de pen-tests client-side.

Client-side es un subgénero de pentest

OSINT automática ya forma parte de los pentests profesionales.

- ▶ **Phishing** and explotación de apls. cliente llevaron al desarrollo de pen-tests client-side.
- ▶ Client-side pentests automatizados incluyen **recolección** abierta de emails, contenido **falso** en mails y urls o archivos elaborados para **explotar** vulns.

Client-side es un subgénero de pentest

OSINT automática ya forma parte de los pentests profesionales.

- ▶ **Phishing** and explotación de apls. cliente llevaron al desarrollo de pen-tests client-side.
- ▶ Client-side pentests automatizados incluyen **recolección** abierta de emails, contenido **falso** en mails y urls o archivos elaborados para **explotar** vulns.
- ▶ Core Security Consulting Services proveyó amablemente estas estadísticas:
 - ▶ 2007: 23 % pentests, tuvieron CS.
 - ▶ 2008: 11 % pentests, tuvieron CS.

Client-side es un subgénero de pentest

OSINT automática ya forma parte de los pentests profesionales.

- ▶ **Phishing** and explotación de apl. cliente llevaron al desarrollo de pen-tests client-side.
- ▶ Client-side pentests automatizados incluyen **recolección** abierta de emails, contenido **falso** en mails y urls o archivos elaborados para **explotar** vulns.
- ▶ Core Security Consulting Services proveyó amablemente estas estadísticas:
 - ▶ 2007: 23 % pentests, tuvieron CS.
 - ▶ 2008: 11 % pentests, tuvieron CS.
- ▶ Luego, involucra **OSINT**, **ingeniería social** y exploits binarios.

(FriendAds: ver charla "Smartphones (in)security.^{en} Ekoparty 2008, es sobre exploits binarios client-side)



Trabajo relacionado

- ▶ Paterva es una pequeña compañía, liderada por Raelof Temmingh, desarrollando un programa de OSINT llamado Maltego [1][2][3].

Trabajo relacionado

- ▶ Paterva es una pequeña compañía, liderada por Raelof Temmingh, desarrollando un programa de OSINT llamado Maltego [1][2][3].
- ▶ Ellos se concentran en **Transforms** (e.g.: domain2emails) y en una especificación de arquitectura cliente-servidor. En nuestro caso llamamos **Expanders** a algo similar a las *Transforms*.

Trabajo relacionado

- ▶ Paterva es una pequeña compañía, liderada por Raelof Temmingh, desarrollando un programa de OSINT llamado Maltego [1][2][3].
- ▶ Ellos se concentran en **Transforms** (e.g.: domain2emails) y en una especificación de arquitectura cliente-servidor. En nuestro caso llamamos **Expanders** a algo similar a las *Transforms*.
- ▶ Ellos usan un rico conjunto de **entidades** como Person, Domain, Host, etc.

Trabajo relacionado

- ▶ Paterva es una pequeña compañía, liderada por Raelof Temmingh, desarrollando un programa de OSINT llamado Maltego [1][2][3].
- ▶ Ellos se concentran en **Transforms** (e.g.: domain2emails) y en una especificación de arquitectura cliente-servidor. En nuestro caso llamamos **Expanders** a algo similar a las *Transforms*.
- ▶ Ellos usan un rico conjunto de **entidades** como Person, Domain, Host, etc.
- ▶ También la librería PyMaltego esta siendo desarrollada por the grugq, incluyendo funcionalidad cliente-servidor.

Que es lo nuevo

- ▶ Recolector y analizador de múltiples redes sociales/cualquier dominio(Exomind):
 - ▶ Módulos (casi) **conectables**, Bots implementando Expanders.
 - ▶ **Fusion** online de alias de personas/nodos durante la recolección.

Que es lo nuevo

- ▶ Recolector y analizador de múltiples redes sociales/cualquier dominio(Exomind):
 - ▶ Módulos (casi) **conectables**, Bots implementando Expanders.
 - ▶ **Fusion** online de alias de personas/nodos durante la recolección.
- ▶ Pesaje online de nodos y enlaces permitiendo **filtrado** y resultados mas precisos.

Que es lo nuevo

- ▶ Recolector y analizador de múltiples redes sociales/cualquier dominio(Exomind):
 - ▶ Módulos (casi) **conectables**, Bots implementando Expanders.
 - ▶ **Fusion** online de alias de personas/nodos durante la recolección.
- ▶ Pesaje online de nodos y enlaces permitiendo **filtrado** y resultados mas precisos.
- ▶ Usando la información juntada:
 - ▶ **Impersonación** de vocabulario para afinar tus **chatbots** (spanglish alert!).
 - ▶ Armazón **programático** para algoritmos de análisis y exportación a otros entornos.

Que es lo nuevo

- ▶ Recolector y analizador de múltiples redes sociales/cualquier dominio(Exomind):
 - ▶ Módulos (casi) **conectables**, Bots implementando Expanders.
 - ▶ **Fusion** online de alias de personas/nodos durante la recolección.
- ▶ Pesaje online de nodos y enlaces permitiendo **filtrado** y resultados mas precisos.
- ▶ Usando la información juntada:
 - ▶ **Impersonación** de vocabulario para afinar tus **chatbots** (spanglish alert!).
 - ▶ Armazón **programático** para algoritmos de análisis y exportación a otros entornos.
- ▶ **Extra!** Chatbot basado en Motor de Búsqueda.



Descripción general de Exomind

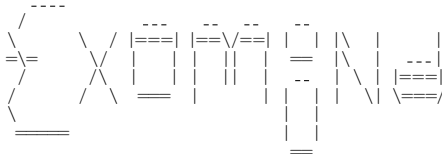
- Clase exomind: la interfaz de líneas de comando Python.

Descripción general de Exomind

- ▶ Clase exomind: la interfaz de líneas de comando Python.
- ▶ Clase Exomind: interfaz **programática**.

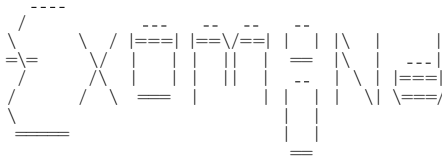
Descripción general de Exomind

- ▶ Clase `exomind`: la interfaz de líneas de comando Python.
- ▶ Clase `Exomind`: interfaz **programática**.
- ▶ Clase `SQLGraph`: abstracción de grafo sobre MySQL, similar a la provista por la librería `networkx` (pero persistente).
 - ▶ Modelo simple de entidades: `node`, `edge`, `node_attr` and `edge_attr`. También los atributos `tiennentype` y `count`.



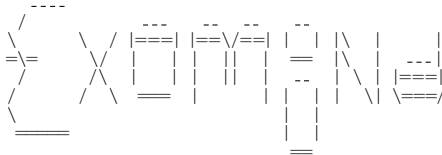
Descripción general de Exomind (cont.)

- ▶ BlackWidow: el recolector/rastreador, soporta concurrencia por lotes, BFS y DFS (usando SQLQueue).



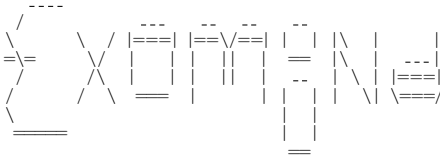
Descripción general de Exomind (cont.)

- ▶ BlackWidow: el recolector/rastreador, soporta concurrencia por lotes, BFS y DFS (usando SQLQueue).
- ▶ LambHerd: inicia los bots.



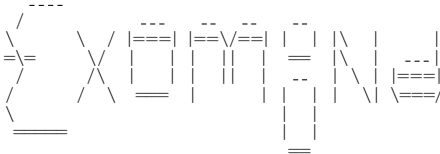
Descripción general de Exomind (cont.)

- ▶ BlackWidow: el recolector/rastreador, soporta concurrencia por lotes, BFS y DFS (usando SQLQueue).
- ▶ LambHerd: inicia los bots.
- ▶ mechanize para guiones web [4], e.g. login .



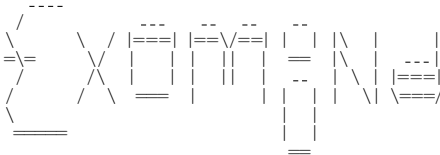
Descripción general de Exomind (cont.)

- ▶ BlackWidow: el recolector/rastreador, soporta concurrencia por lotes, BFS y DFS (usando SQLQueue).
- ▶ LambHerd: inicia los bots.
- ▶ mechanize para guiones web [4], e.g. login .
- ▶ pygraphviz+networkx para visualización [5].



Descripción general de Exomind (cont.)

- ▶ BlackWidow: el recolector/rastreador, soporta concurrencia por lotes, BFS y DFS (usando SQLQueue).
- ▶ LambHerd: inicia los bots.
- ▶ mechanize para guiones web [4], e.g. login .
- ▶ pygraphviz+networkx para visualización [5].
- ▶ msnlib para el ejemplo de IM chatbot [6].



Modulos Bots

Recuperan y adaptan información pero también pueden ser interactivos:

- ▶ **Servicios** de redes sociales o cualq. tipo.
- ▶ Info general de Motores de **Búsqueda**.



Modulos Bots

Recuperan y adaptan información pero también pueden ser interactivos:

- ▶ **Servicios** de redes sociales o cualq. tipo.
- ▶ Info general de Motores de **Búsqueda**.
- ▶ Algoritmos de grafos para procesar información existente.



Métodos de los Bots

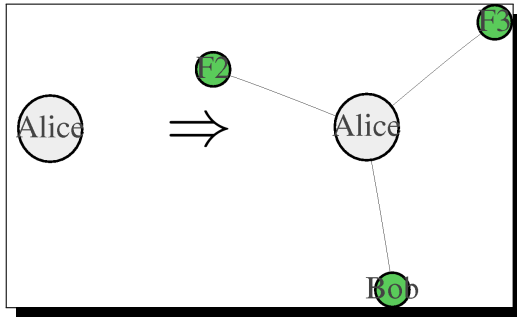
Los Bots deben tener (algunas) de las siguientes habilidades:

- ▶ **Expanders:** devuelven vecinos/contactos de un nodo (más atributos).
- ▶ **Weigh Scale:** mide el peso de un nodo o enlace y filtra.
- ▶ Comandos **ChatBot:** contactar personas y de manera semi/compl. automática hablar con ellos.



Expander

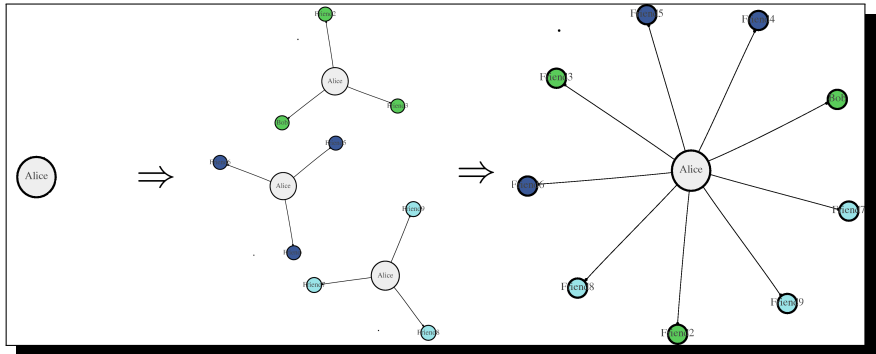
Por ejemplo, para el servicio ViralBuddy se puede implementar el ViralBuddyBot que expande a Alice:



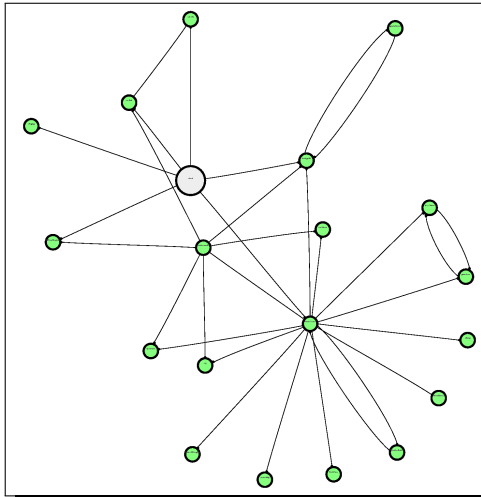
Más extra **atributos** para Alice, cada amigo y cada enlace por separado.

Fusión de múltiples redes sociales

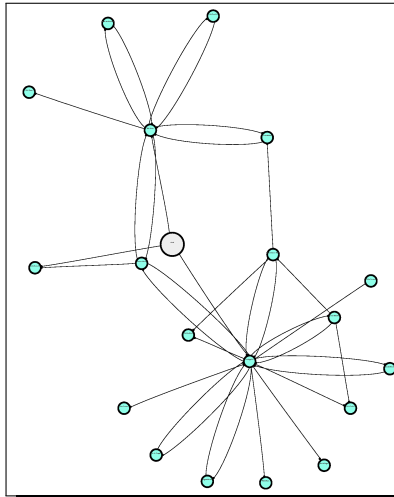
Los contactos de Alice de varias redes sociales pueden ser cruzados:



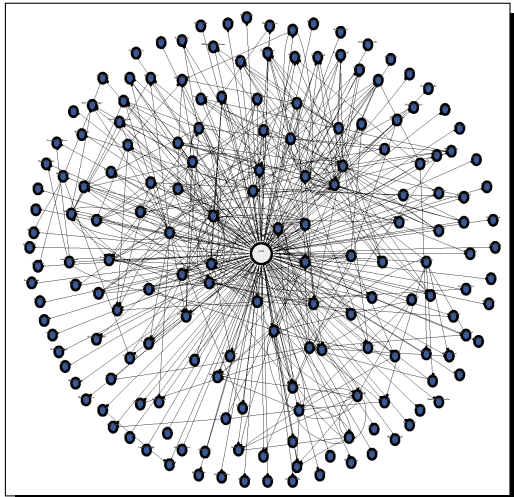
Twitter Following



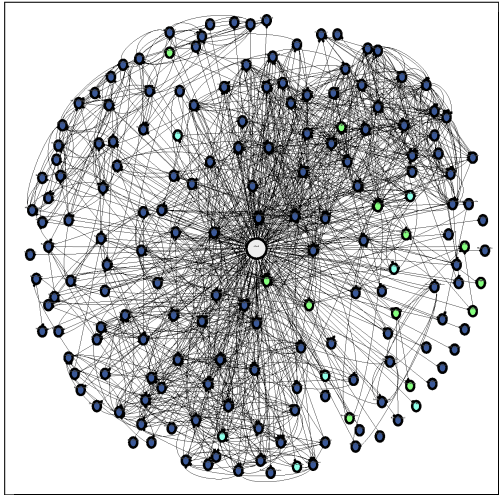
LinkedIn Recommendations



Facebook Friends



Cuando se recorren juntas...



Ejemplos

```
>>> all_expanders()  
LinkedInBot::recommendations  
LinkedInBot::also_viewed  
GraphBot::neighbors  
GraphBot::with_all  
SearchEngineBot::domain_to_emails  
SearchEngineBot::name_to_self_emails  
SearchEngineBot::name_to_emails_strong  
SearchEngineBot::domain_to_emails_strong  
SearchEngineBot::name_to_emails  
SearchEngineBot::vocabulary  
FacebookBot::friends  
TwitterBot::following  
TwitterBot::followers  
TwitterBot::favorites
```

Evaluación recursiva de regexs

Información general y **desestructurada** proveniente de **resultados de motor de búsqueda** (SearchEngineBot) pueden ser recuperados y procesados con fines **específicos**.

- ▶ Teléfonos.
- ▶ Direcciones físicas.
- ▶ E-mails (**prototipo**).

comments. Main Content. English Wikipedia references for Coresecurity.com 1—10 of 10 ... From: CORE Security Technologies Advisories *advisories@coresecurity.com* ... Cor Security Technologies Advisory <http://www.coresecurity.com> Axis Network ... Gerardo Richarte and CoreLabs. *gera@coresecurity.com*. MD5 to be considered \xa1Gracias! gera (Gerardo Richarte) *gerardo.richarte@coresecurity.com* or http://oss.coresecurity.com/uhooker/release/1.2/uhooker_v1.2.zip (zip) checkout the doc pages because Im constantly posting new stuff like sample scripts ... Advisory URL: <http://www.coresecurity.com/?action=item&id=2035> ... <http://www.coresecurity.com/files/attachments/CORE-2007-1004-VLC-tutoria> ... Browser Defender site report for coresecurity



Evaluación recursiva de regexs (cont.)

Para cosechar la data específica de los snippets or páginas completas necesitas poderosas (y *muchas!*) regular expressions para relacionar la data.

- E-mail: firstlastname usuario@domain.com. E.g., si tienes el usuario:

```
tokregex+'_'+tokregex+'_'+user+'_at_'+tokregex
```

Evaluación recursiva de regexs (cont.)

Para cosechar la data específica de los snippets or páginas completas necesitas poderosas (y *muchas!*) regular expressions para relacionar la data.

- ▶ E-mail: firstlastname usuario@domain.com. E.g., si tienes el usuario:

```
tokregex+'_' + tokregex+'_' + user+'_at_' + tokregex
```

- ▶ Muchas heurísticas para post-procesar la coincidencias de las regexs.

```
Alice Smith a.....@bla.com —>
```

```
Alice Smith asmith@bla.com
```

```
_ _ alice.smith@bla.com —>
```

```
Alice Smith alice.smith@bla.com
```



Evaluación recursiva de regexs (cont.)

Para cosechar la data específica de los snippets or páginas completas necesitas poderosas (y *muchas!*) regular expressions para relacionar la data.

- ▶ E-mail: firstlastname usuario@domain.com. E.g., si tienes el usuario:

```
tokregex+'_' + tokregex+'_' + user+'_at_' + tokregex
```

- ▶ Muchas heurísticas para post-procesar la coincidencias de las regexs.

```
Alice Smith a.....@bla.com —>
```

```
Alice Smith asmith@bla.com
```

```
_ _ alice.smith@bla.com —>
```

```
Alice Smith alice.smith@bla.com
```

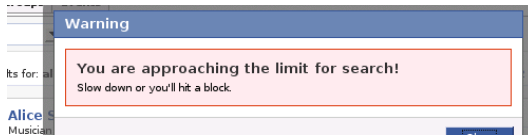
- ▶ **Iterar**, si hay cercanía asumir **relación** y definir un Expander!



Previnendo abusos y prohibiciones

El **sigilo** es una preocupación fundamental de los pentesters profesionales, luego para reducir el ruido generado todos los bots incluyen:

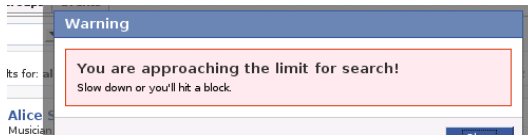
- ▶ `sleep_regular_secs`: dormir t segundos cuando es necesario.



Previnendo abusos y prohibiciones

El **sigilo** es una preocupación fundamental de los pentesters profesionales, luego para reducir el ruido generado todos los bots incluyen:

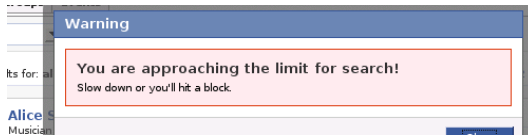
- ▶ `sleep_regular_secs`: dormir t segundos cuando es necesario.
- ▶ `sleep_random_bool`: dormir tiempo $2 * t * rand(0, 1)$ entre ops. si se habilita.



Previnendo abusos y prohibiciones

El **sigilo** es una preocupación fundamental de los pentesters profesionales, luego para reducir el ruido generado todos los bots incluyen:

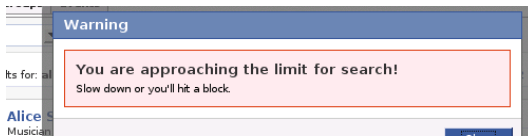
- ▶ `sleep_regular_secs`: dormir t segundos cuando es necesario.
- ▶ `sleep_random_bool`: dormir tiempo $2 * t * rand(0, 1)$ entre ops. si se habilita.
- ▶ `sleep_module_gets`: dormir luego de esta cantidad de ops.



Previnendo abusos y prohibiciones

El **sigilo** es una preocupación fundamental de los pentesters profesionales, luego para reducir el ruido generado todos los bots incluyen:

- ▶ `sleep_regular_secs`: dormir t segundos cuando es necesario.
- ▶ `sleep_random_bool`: dormir tiempo $2 * t * rand(0, 1)$ entre ops. si se habilita.
- ▶ `sleep_module_gets`: dormir luego de esta cantidad de ops.
- ▶ La línea famosa `Press Enter to continue`: puede ser incluida en tus bots para incorporar interacción humana (y frenar!).



Demo1: rastreo dirigido de emails!

LeakedOut: las redes sociales que te atrapan

Pesaje y filtrado

Tenemos ahora mucha info, pero es debil:

- ▶ Nombres o nicks **comunes** **polucionan** la información de la **nube**, e.g. Pablo and Juan .
- ▶ **Poco/ningun contexto**: la información es **ambigua**.



Pesaje y filtrado

Tenemos ahora mucha info, pero es debil:

- ▶ Nombres o nicks **comunes** **polucionan** la información de la **nube**, e.g. Pablo and Juan .
- ▶ **Poco/ningun contexto**: la información es **ambigua**.
- ▶ Detectar **alias** y **duplicados** es un problema complejo.



Pesaje y filtrado

Tenemos ahora mucha info, pero es debil:

- ▶ Nombres o nicks **comunes** **polucionan** la información de la **nube**, e.g. Pablo and Juan .
- ▶ **Poco/ningun contexto**: la información es **ambigua**.
- ▶ Detectar **alias** y **duplicados** es un problema complejo.
- ▶ Pero por lo menos podemos **filtrar**, solo información **confirmable** y **precisa** será permitida.



Pesaje y filtrado

Tenemos ahora mucha info, pero es debil:

- ▶ Nombres o nicks **comunes** **polucionan** la información de la **nube**, e.g. Pablo and Juan .
- ▶ **Poco/ningun contexto**: la información es **ambigua**.
- ▶ Detectar **alias** y **duplicados** es un problema complejo.
- ▶ Pero por lo menos podemos **filtrar**, solo información **confirmable** y **precisa** será permitida.
- ▶ Como? Usando un corpus de referencia universal, algún gran **motor de búsqueda**!



Hits en Motores de Búsqueda

Si estoy impersonando a Alice Smith, es el contacto Bob Johnson un buen **objetivo**?

- ▶ Si Bob tiene un nombre muy comun: **NO**.
- ▶ Si Bob no esta muy relacionado a Alice: **NO**.

$$h(\text{alice smith}) = 289,000$$

Results 1 - 10 of about 289,000 for "alice [smith](#)". (0.38 seconds)

$$h(\text{bob johnson}) = 861,000$$

Results 1 - 10 of about 861,000 for "[bob johnson](#)". (0.20 seconds)

$$h(\text{alice smith,bob johnson}) = 825$$

Results 1 - 10 of about 825 for "alice [smith](#)" "[bob johnson](#)". (0.28 seconds)

Hits en Motores de Búsqueda (cont.)

Normalizar el # de hits (h) y transformarlos en entropía (!):

- ▶ Páginas totales: $M = 21,910,000,000$

Results **1 - 10** of about **21,910,000,000** for **a.** (0.13 seconds)

Hits en Motores de Búsqueda (cont.)

Normalizar el # de hits (h) y transformarlos en entropía (!):

- ▶ Páginas totales: $M = 21,910,000,000$

Results 1 - 10 of about 21,910,000,000 for a. (0.13 seconds)

- ▶ Entropía máx. (1 hit): $-\log_2 M \simeq 34,35$ bits

Hits en Motores de Búsqueda (cont.)

Normalizar el # de hits (h) y transformarlos en entropía (!):

- ▶ Páginas totales: $M = 21,910,000,000$

Results 1 - 10 of about 21,910,000,000 for a. (0.13 seconds)

- ▶ Entropía máx. (1 hit): $-\log_2 M \simeq 34,35$ bits
- ▶ Fraction de Bob': $\frac{861,000}{21,910,000,000} \simeq 3,929 \cdot 10^{-5}$

Hits en Motores de Búsqueda (cont.)

Normalizar el # de hits (h) y transformarlos en entropía (!):

- ▶ Páginas totales: $M = 21,910,000,000$

Results 1 - 10 of about **21,910,000,000** for a. (0.13 seconds)

- ▶ Entropía máx. (1 hit): $-\log_2 M \simeq 34,35$ bits
- ▶ Fraction de Bob': $\frac{861,000}{21,910,000,000} \simeq 3,929 \cdot 10^{-5}$
- ▶ Entropía: $-\log_2(3,929 \cdot 10^{-5}) \simeq 14,635$ bits

Hits en Motores de Búsqueda (cont.)

Normalizar el # de hits (h) y transformarlos en entropía (!):

- ▶ Páginas totales: $M = 21,910,000,000$

Results 1 - 10 of about **21,910,000,000** for a. (0.13 seconds)

- ▶ Entropía máx. (1 hit): $-\log_2 M \simeq 34,35$ bits
- ▶ Fraction de Bob': $\frac{861,000}{21,910,000,000} \simeq 3,929 \cdot 10^{-5}$
- ▶ Entropía: $-\log_2(3,929 \cdot 10^{-5}) \simeq 14,635$ bits
- ▶ Normalizado (betw. 0 and 1): $\frac{34,35-14,635}{34,35} \simeq 0,573$

Hits en Motores de Búsqueda (cont.)

Normalizar el # de hits (h) y transformarlos en entropía (!):

- ▶ Páginas totales: $M = 21,910,000,000$

Results 1 - 10 of about **21,910,000,000** for a. (0.13 seconds)

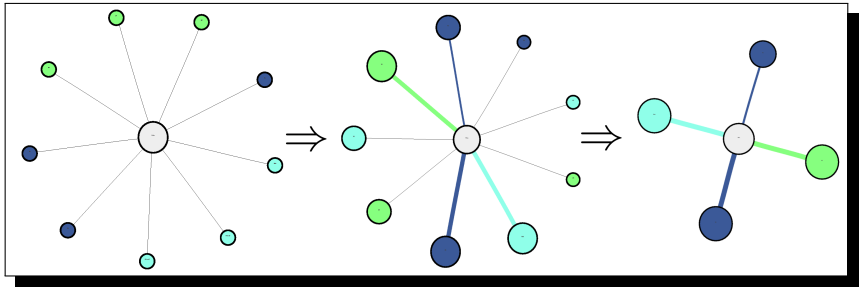
- ▶ Entropía máx. (1 hit): $-\log_2 M \simeq 34,35$ bits
- ▶ Fraction de Bob': $\frac{861,000}{21,910,000,000} \simeq 3,929 \cdot 10^{-5}$
- ▶ Entropía: $-\log_2(3,929 \cdot 10^{-5}) \simeq 14,635$ bits
- ▶ Normalizado (betw. 0 and 1): $\frac{34,35-14,635}{34,35} \simeq 0,573$
- ▶ Si por ejemplo 0,5 es el **límite** elegido, Bob es **descartado**!

Distancia Normalizada de Google (NGD)

Lo mismo con **conexiones**. Hits $h(x)$ and $h(x, y)$ son mas utiles si se normalizan (donde $p(x) = \frac{h(x)}{M}$ y $p(x|y) = \frac{p(x,y)}{p(y)}$):

- ▶ Distancia de Jaccard: $\frac{h(x,y)}{h(x)+h(y)-h(x,y)}$
- ▶ *NGD* es más robusta! [7]:
$$\frac{\max\{-\log_2 p(x|y), -\log_2 p(y|x)\}}{\max\{-\log_2 p(x), -\log_2 p(y)\}} = \frac{\max\{\log_2 h(x), \log_2 h(y)\} - \log_2 h(x,y)}{\max\{-\log_2 h(x), -\log_2 h(y)\}}$$
- ▶ $NGD(\text{Alice Smith}, \text{Bob Johnson}) \simeq 0,545$ si, por ejemplo, 0,5 es nuestro **límite** elegido, la conexión con Bob es **descartada**, no es un objetivo fiable para contactar!

Resultado del pesaje y filtrado



Observación: Un **contexto** puede ser elegido para restringir el corpus! E.g. security, argentina. etc.

Weigh Scales de Exomind

```
>>> all_weigh_scales()  
SearchEngineBot::jaccard_distance  
SearchEngineBot::se_hits  
SearchEngineBot::normalized_se_entropy  
SearchEngineBot::normalized_se_distance  
SearchEngineBot::hits_distance  
GraphBot::obfuscate  
>>>
```

Las *Weigh scales* pueden aplicar transformaciones a los entornos, por ejemplo `GraphBot::obfuscate` **encripta** los datos recolectados. También se pueden detectar **ciegamente** conexiones privadas filtradas...



Demo2: Reconstrucción ciega de redes! Fuerza bruta de todos los posibles links!

Usando la Data Recolectada

- ▶ Cualquier tipo de info disponible puede ser recolectada. Solamente hay que añadir un nuevo `Attribute` y extraerlo dentro de un viejo o nuevo *expander*.

Usando la Data Recolectada

- ▶ Cualquier tipo de info disponible puede ser recolectada. Solamente hay que añadir un nuevo `Attribute` y extraerlo dentro de un viejo o nuevo *expander*.
- ▶ Un atributo especial llamado TAG esta diseñado para construir **bolsas de palabras** definiendo **perfiles difusos** para personas.

Usando la Data Recolectada

- ▶ Cualquier tipo de info disponible puede ser recolectada. Solamente hay que añadir un nuevo Attribute y extraerlo dentro de un viejo o nuevo *expander*.
- ▶ Un atributo especial llamado TAG esta diseñado para construir **bolsas de palabras** definiendo **perfiles difusos** para personas.

“ He is the best reverse
engineer and security...

```
add_node_attr('Bob_Johnson', Attributes.TAG, 'best')  
add_node_attr('Bob_Johnson', Attributes.TAG, 'reverse')  
add_node_attr('Bob_Johnson', Attributes.TAG, 'engineer')  
add_node_attr('Bob_Johnson', Attributes.TAG, 'security')
```



Impersonación de vocabulario

Si el perfil difuso esta centrado en texto escrito del nodo objetivo, estamos recolectando su vocabulario. Usamos patrones comunes:

- ▶ ...Alice said...,
- ▶ ...posted by Alice...,
- ▶ ...Alice wrote....

```
list: [('co-founder', 6L), ('sets', 7L), ('technical', 8L), ('direction', 6L),  
( 'company', 17L), ('responsible', 5L), ('overseeing', 5L), ('development', 3L),  
( 'managed', 4L), ('catch', 4L), ('chief', 43L), ('technology', 3L), ('officer', 40L),  
( 'talk', 4L), ('new', 8L), ('class', 3L), ('vulnerability', 15L), ('thats', 2L),  
( 'talks', 2L), ('recent', 3L), ('update', 2L), ('interview', 6L), ('said', 98L), ('view',  
5L), ('information', 4L), ('provided', 3L), ('free', 4L), ('public', 8L), ('search', 5L),  
( 'listing', 4L), ('friends', 5L), ('photos', 6L), ('videos', 6L), ('join', 5L),  
( 'attacker', 2L), ('remotely', 1L), ('execute', 1L), ('code', 12L), ('exploit', 9L),  
( 'bugs', 10L), ('user', 4L), ('interaction', 3L), ('releasing', 1L), ('customers', 1L),  
( 'week', 7L), ('technology', 39L), ('flaw', 9L), ('demonstrates', 1L), ('writes', 1L),  
( 'discusses', 1L), ('security', 37L), ('threats', 2L), ('tools', 4L), ('importance', 1L),  
( 'presentation', 3L), ('titled', 2L), ('esteemed', 2L), ('reply', ...)
```

Impersonación de vocabulario (cont.)

(Diversión previa con lenguajes [8]).

- Modulo llamado *Dino*, basado en un **thesaurus** y con falta de chequeos de ambigüedad en este momento ;(, hackers de PLN se buscan.

Bob: tell me

(Falsa Alice: *tell* me that)

Luego de la traducción de vocabulario!

Falsa Alice: **state** me that

Bob: tell me

(Falsa Alice: *do* your friends *put* money in your *pocket*?)

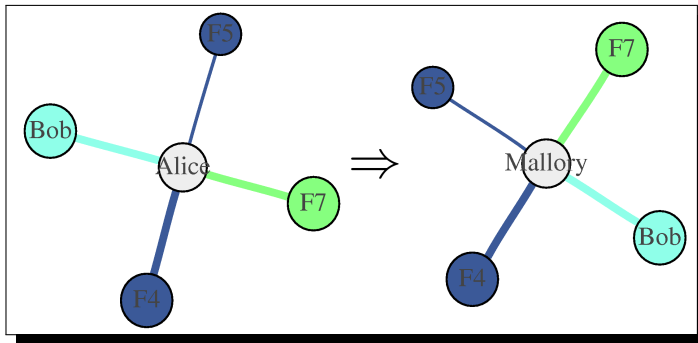
Traducir!

Falsa Alice: **execute** your friends **risk** money in your **collect**



Infiltración de red social

Mallory quiere automáticamente impersonar a Alice e interactuar con su contacto Bob, posiblemente aprovechando credenciales robadas. Cualquier canal de comunicación.



Infiltración de red social (cont.)

Em Exomind se incluyó una infiltración de chatbot MSN:

- ▶ Mallory usa **nick** de Alice .
- ▶ Posibles usuarios MSN son recuperados de los contactos de Alice en la red recolectada.
- ▶ Mensajes iniciales se envían cada N segundos.
- ▶ Respuestas automáticas o manuales usando el Chatbot de películas o Eliza, y el **vocabulario** recolectado.

Chatbot de Motor de Búsqueda

Que diría Google (Search) si pudiera hablar? Sistemas de respuestas a preguntas existen ya [9] pero queremos una charla **informal**.

Guiones de películas: formato mas o menos estructurado [10].

- Consultar chatline or chatline[: -1] hasta encontrar respuestas.

(También se puede usar Google Translate para cambiar de idioma.)



Chatbot de Motor de Búsqueda

Que diría Google (Search) si pudiera hablar? Sistemas de respuestas a preguntas existen ya [9] pero queremos una charla **informal**.

Guiones de películas: formato mas o menos estructurado [10].

- ▶ Consultar chatline or chatline[: -1] hasta encontrar respuestas.
- ▶ Extraer todas las posibles respuestas de los **snippets**.

(También se puede usar Google Translate para cambiar de idioma.)

Chatbot de Motor de Búsqueda

Que diría Google (Search) si pudiera hablar? Sistemas de respuestas a preguntas existen ya [9] pero queremos una charla **informal**.

Guiones de películas: formato mas o menos estructurado [10].

- ▶ Consultar chatline or chatline[: -1] hasta encontrar respuestas.
- ▶ Extraer todas las posibles respuestas de los **snippets**.
- ▶ Elegir un subconjunto al **azar**, digamos de 20.

(También se puede usar Google Translate para cambiar de idioma.)

Chatbot de Motor de Búsqueda

Que diría Google (Search) si pudiera hablar? Sistemas de respuestas a preguntas existen ya [9] pero queremos una charla **informal**.

Guiones de películas: formato mas o menos estructurado [10].

- ▶ Consultar chatline or chatline[: -1] hasta encontrar respuestas.
- ▶ Extraer todas las posibles respuestas de los **snippets**.
- ▶ Elegir un subconjunto al **azar**, digamos de 20.
- ▶ (opcional!) Medir la **distancia** de Google y elegir las mejores 5 respuestas.

(También se puede usar Google Translate para cambiar de idioma.)

Chatbot de Motor de Búsqueda

Que diría Google (Search) si pudiera hablar? Sistemas de respuestas a preguntas existen ya [9] pero queremos una charla **informal**.

Guiones de películas: formato mas o menos estructurado [10].

- ▶ Consultar chatline or chatline[: -1] hasta encontrar respuestas.
- ▶ Extraer todas las posibles respuestas de los **snippets**.
- ▶ Elegir un subconjunto al **azar**, digamos de 20.
- ▶ (opcional!) Medir la **distancia** de Google y elegir las mejores 5 respuestas.
- ▶ Elegir la respuesta mas usada del subconjunto.

(También se puede usar Google Translate para cambiar de idioma.)

Chatbot de Motor de Búsqueda

Que diría Google (Search) si pudiera hablar? Sistemas de respuestas a preguntas existen ya [9] pero queremos una charla **informal**.

Guiones de películas: formato mas o menos estructurado [10].

- ▶ Consultar chatline or chatline[: -1] hasta encontrar respuestas.
- ▶ Extraer todas las posibles respuestas de los **snippets**.
- ▶ Elegir un subconjunto al **azar**, digamos de 20.
- ▶ (opcional!) Medir la **distancia** de Google y elegir las mejores 5 respuestas.
- ▶ Elegir la respuesta mas usada del subconjunto.
- ▶ Voilà!

(También se puede usar Google Translate para cambiar de idioma.)

Demo3: Chatbot de Películas basado en Motor de Búsqueda!



Conclusion/Countermeasures

- ▶ La ciudad es un pañuelo.
- ▶ No usar tu verdadero nombre en la web (en la **economía de la atención** la información es moneda, no la regales!).
- ▶ No aceptes extraños en tu red social (en la **economía social** los contactos son moneda, no entregues a tus amigos o ti mismo!).
- ▶ Usar mensajería interna para temas privados de tu organización.
- ▶ Analiza tus redes sociales con Exomind! [11]

Preguntas Finales?

Gracias!

Aure: discusiones de webapps y python.

Alfred: plantilla LaTeX.

Beto: Estads. de consultoría.

ITBA team: entorno científico.

Core team: comida, cobijo y entorno creativo de hacking.



Referencias

- 1 www.paterva.com/maltego
- 2 www.first.org/conference/2007/papers/temmingh-roelof-slides.pdf
- 3 www.blackhat.com/presentations/bh-europe-08/Temmingh-Bohme/Presentation/bh-eu-08-temmingh-bohme.pdf
- 4 <http://wwwsearch.sourceforge.net/mechanize/>
- 5 <https://networkx.lanl.gov/wiki/pygraphviz/>
- 6 <http://blitiri.com.ar/p/msnlib/>
- 7 R. Cilibrasi, P.M.B. Vitanyi, Automatic meaning discovery using Google.
<http://xxx.lanl.gov/abs/cs.CV/0312044> (2004)
- 8 Look Who's Translating: Impersonations, Chinese Whispers and Fun with Machine Translation on the Internet www.mt-archive.info/EAMT-2006-Gaspari.pdf
- 9 Quarteroni et. al, A Chatbot-based Interactive Question Answering System
- 10 www.imsdb.com
- 11 Herramienta Exomind, descargas pronto en corelabs.coresecurity.com