

Evolution

or...I know what you (and your company) did last summer

CanSecWest 2007



A NEW TRAIN OF THOUGHT

PATERVA

Agenda



- Introduction
- Things are changing...
- Part I – Finding info
- Demo
- Interlude
- Part II – uses of the information collected
- Conclusion

Introduction



- Who I am?
 - Roelof Temmingh
 - Completed Bachelor's degree in Electronic Engineering
 - Started SensePost with friends in 2000
 - Talked/trained at BlackHat, RSA, Defcon, FIRST etc.
 - Co-wrote some Syngress books
 - Built tools @ SensePost – Wikto, Suru, BiDiBLAH, Crowbar
- Today
 - Started Paterva beginning of 2007
 - R&D

Think deep...introspection

Why do we hack?

No...I mean...really...

“See if it's safe because I bank there” - yeah...OK if you say so.

The pleasure from seeing the lock open == the info appearing on your screen. Control over the lock.

It's being able to do things normal people can't, knowing things that normal people don't.

It's less about applying the information->knowledge

Things are changing (at least from the outside)



- When last did you fully own a box? (root/system)
- When last did you fully own a database?
- When was the last proper worm?
- Where have all the script kiddies gone??

- At least from outside – you are dealing with 3 ports – 80,443 and 25.
- If you own something today it's probably via a web application gone wrong, or it's a client via a vuln in a client side app (think browser, email client).

Job security - NOT!



Company X asks for an external security assessment.

Company X has one static website (and email)

As a security expert you need to assess it.

- Nmap, Nessus, Wikto/Nikto – now what?
- The most used tab in Wikto?

Are they really worried about an 0day on their IIS6?

Or that 'bad things happen on the Internet' ?

The unknown almost always scares..

Hacker's 6th sense

People regularly ask me - what can you find out about...?

... without touching/hacking them

... without them knowing

Most of the time it turns out there is a lot to be found

“Why couldn't you do this??”

- Because we know the fabric of the 'net
- Because we think different thoughts
- Because we are have built-in deviousness
- Because we know who and how to ask

“But why would you like to know this?”

Hackers are not interested in the application of data, but know how to get to it.

People always ask me...



Non-tech people ask “what can you tell me about this:

- Email address
- Company
- Person
- Word/phrase
- Website

...and us tech people think:

- IP addresses, virtual hosts
- Netblocks / AS routes
- Affiliations (Linkedin, Myspace, Orkut, Zoominfo, Facebook etc)
- Microformats - XFN, vCards, hCards etc. etc.
- Forward and reverse DNS – MX/NS records
- Whois records / rWhois and referring registrars
- Google fu / Deep web search
- Services like ip2geo, wayback machine, Google Earth, Zoominfo
- Meta information in documents

Part I – relationship collection

The thinking behind the framework



Step 1: What can you tell me about...

- How would you do it as a human? =>
- Information transforms in the framework

Step 2: Collect information

- Searching, surfing
- Deep web & services

Step 3: Parse/convert information to **new entities**

Step 4: **Goto 1**

Find the hidden information -

A -> B -> C and

X -> Y -> C

..then $A \approx X$

Part I – relationship collection

Thinking behind the framework



- Entity or Entities -> **Transform** -> Entity or Entities
- Repeat

Entities I have now are:

```
core/entities roeloftemmingh$ ls -l
```

- AffiliationEntity.java
- DNSNameEntity.java
- DocumentEntity.java
- DomainEntity.java
- EmailAddressEntity.java
- IPAddressEntity.java
- LocationEntity.java
- PersonEntity.java
- PhraseEntity.java
- TelephoneNumberEntity.java
- WebSiteEntity.java

Part I – relationship collection: Transforms



Transforms are: (as of 3 April 2007):

```
/core/transforms roelofteemingh$ ls -l  
DNSNameToDomain.java  
DNSNameToIPAddress.java  
DNSNameToWebSite.java  
DomainToDNSNameBrute.java  
DomainToDNSNameMX.java  
DomainToDocument.java  
DomainToEmailPhoneSiteGoogle.java  
DomainToEmailWhois.java  
DomainToTelephoneWhois.java  
EmailToAffMySpace.java  
EmailToDomain.java  
EmailToEmailPhoneSiteGoogle.java  
IPAddressToDNSName.java  
IPAddressToDNSNameVH.java  
IPAddressToEmailPhoneWhois.java  
IPAddressToLocation.java  
PersonToAffLinkedIn.java  
PersonToEmailPhoneSiteGoogle.java  
PersonToEmailPhoneSiteGoogleBlog.java  
PhraseToEmailPhoneSiteGoogle.java  
PhraseToEmailPhoneSiteGoogleBlog.java  
TelephoneToEmailPhoneSiteGoogle.java  
WebSiteToDNSName.java  
WebSiteToWebSiteIncomingLink.java  
_WebSiteGetFirstSeen.java  
_WebSiteGetServerVersion.java
```

26 transforms
and growing steadily..

Part I – relationship collection

The thinking behind the framework



- Transforms - “who can do anything with a ...?”
- Entities open and expandable
- Transforms uses plugin architecture
- Thus, you are only limited by your own imagination
- If you are in the Lexis/Nexis club...
- If you work for the phone company
- If you have access to any other juicy databases etc.

Transforms



Some really easy examples

- DNS name -> IP number(s)
- IP number -> DNS name
- Domain (whois) -> email address(es)
- IP address (whois) -> telephone number
- Telephone number -> Geo location
- IP number -> Geo location

Transforms tel -> email



Some more interesting examples:

Consider Telephone number -> email address
How would a human do it?

- Step 1: Google the telephone number
- Step 2: Look at the snippet/Surf to the result
- Step 2: Look for email addresses
- Step 3: See which are clearly connected to the telephone number

But it gets nasty...

Transforms: tel- > email

Assume the number is +27 83 448 6996

- =~ 083 448 6996
- =~ (0)83 448 6996
- =~ 083 4486996
- =~ +2783 4486996
- =~ etc

Option 1: only look for 448 6996.

- Your friendly plumber in Burundi

Option 2: Try combinations

- But “+27 83 448 6996” is more likely to be correct than 083 448 6996

Transforms – confidence levels



Same goes for First Name/Last Name

- Results on search query for “Roelof Temmingh” is more likely to contain the correct email address for me than a query for R Temmingh.

So – let's do all the queries, but give each a 'confidence index'

Are there other parameters we can use to increase the quality of our results?

Transforms: factors when sorting by relevance



- Frequency of the parsed result
If, after parsing, I get roelof.temmingh@gmail.com 100 times it's likely to be more relevant than something that appears 2 times.
- Significance of the site where the term is located
Especially necessary when working with phrases.
- Correlation to the original search term.
roelof.temmingh@gmail.com is more likely to be Roelof Temmingh's email address than kosie.kramer@yahoo.com.
- Proximity to the search term.

Transforms : using Google

So, for each “fuzzy” search (where relationships are not 1:1) I can create an confidence index.

Using Google and only looking at snippets I get:

- Speedy results
- Control over amount of results returned
- Country specific results
- Significance of site -> Page Rank
- Proximity -> Anything in the snippet is already in close proximity to the search term

Confidence index=

(Query confidence * Frequency across all pages * correlation to search term) + (sum of page ranks)

Google helps even more

Numrange – for finding telephone numbers in a certain area.

Applications a.k.a So what??

For conventional security:

- Stock standard footprinting (DNS, IPs, domains etc)
- Nice for finding war dialing (?!) ranges
- Targets for social engineering and client side attacks
- Alternative email addresses for content attacks
 - “if the attacker can get the victim to visit...etc..”
- Finding alliances with weaker security postures
- Understanding business drivers and sensitivities

Applications a.k.a So what??

- Is abc.com a phishing site?

Domain -> email addresses at domain

- > telephone numbers 'connected'

- > related DNS names -> IPs

- > Website

- > Whois -> email addresses

- > whois -> telephone numbers

Telephone numbers -> Geo location(s)

Telephone numbers -> Alternative emails

Telephone numbers -> Related tel. numbers

IPs -> Geo location(s)

IPs -> co-hosted websites on same IP

IPs -> whois -> owners / tel / email

Website -> First seen date

Website -> Phrases

Applications : more interesting stuff



- Who at the NSA uses Gmail?

Domain -> telephone numbers

Telephone number (area) -> Email addresses

- Which NASA employees are using Myspace/LinkedIn?

Domain -> email addresses

Email addresses -> social network

- Which people in Kabul are using Skype?

Telephone number (area) -> email addresses

Email addresses -> Affiliation

Even more applications

- In which countries do the USMC have bases in?

Domain -> Sub domains

Sub domains -> DNS names

DNS names -> IP addresses

IP addresses -> Geo location

Other uses even...

- What are the names and email addresses of single, young woman in my neighbourhood who are straight (or not)?

phone area -> email

email -> social myspace

[filter]

OK..so he's 49 and into pot



1133 West Hastings Street
Vancouver, British Columbia
V6E3T3 Canada

Phone: 1 604 689 9211

Fax: 1 604 689 4358

Sales: 1 604 691 2756

Sales fax: 1 604 691 2791

Toll-free: 1 800 905 8582

Located on the waterfront in
Vancouver Hotel Harbourside

photo tour

leonfia@hotmail.com:200:1lar2lrlgg5pc generator:Goog
luismoran@shaw.ca:200:1cvpro3uvx94y generator:Google
marc@cannabisculture.com:200:ejxpahutkd8a generator:
mary.rideout@fourseasons.com:200:1b2gxj8zzkhf7 gener
pfm@pfmsearch.com:200:5tpk00nmcfd4 generator:GoogleF
slgl@vrrir.de:200:178dts3ex8v generator:GoogleFromT
sreid@pgi.com:200:9o2meafwkxe generator:GoogleFromTe
abgs@mail@abgs.com:100:1spshsjybs4bz generator:Google
aca_online@mac.com:100:12g3ar9wr9i generator:Googl
admin@mascalldance.ca:100:1iaba6xbwz9w1 generator:Gc
sharriott@ngstaff.com:100:1o2rxdpxau7mx generator:Gc

The First Arab-Israeli Joint Conference on Peace and Drugs Policy

If you would like information about this conference, please contact Marc Emery. Phone: (604)

689-0590 Email: Marc@cannabisculture.com ...

www.cannabisculture.com/articles/4815.html - 21k - [Cached](#) - [Similar pages](#)

Marc Emery



View My: [Pics](#) | [Videos](#)

"America's Most
Wanted: Marc Emery,
Prince of Pot"

Male
49 years old
Vancouver, British
Columbia
Canada

Marc Emery's Interests

General

Individual freedom, , marijuana,
women, sex, Video technology,
NOVEMBER 7, 2006 USA election!



Demo
<http://www.paterva.com>

Hold your horses ...a.k.a 'but this is BS'



- “...my mother/grandma can't even operate a mouse...”
- Not everyone on the net gives out their info
- Not all systems readily give info to computers

The info is kept closer...but also wider..

- Whois info details on new domains are now removed
- But look at Myspace / LinkedIn
- Web 2.0 ... new generation of 'WHAT privacy??'
- “Nowadays the kids have their photos on flickr, their profiles and friends on Facebook or Myspace, and their personal thoughts on LiveJournal/Blogger/Twitter, and future-cringeworthy-moments be damned. Concern about privacy is so last century. :o)”

What this “Deep Web” thing anyhow?

Interlude



You are :

- the information you publish
- the information others publish about you
- your associations, and
- **the information you search for.**

The holy grail of information collection – your search terms

How can we know what other people are searching for?

Collecting your search terms



Recently AOL 'lost' a couple of search terms (well OK 20 million of them)

You can search search terms at data.aolsearchlogs.com.

Searches are connected by userID. People are really weird - hours of fun...

- User #13324924 searched on 'help navada a resident of yours is harrassing me and threatening to have me killed by a guy named ronnie and she will not give me my belongings she tookk them and refuses to return my nesseceties' at 2006-04-29 20:11:47
- User #13324924 searched on 'cheap flights' at 2006-05-16 05:43:55

Collecting your search terms



If we control the infrastructure of a network we can

- Redirect outgoing traffic to port 80 into a Squid proxy
- Change two settings in the default Squid config to log MIME headers (cookie) and show parameters (search terms)
- Copy logs into a database
- Extract search terms per Google cookie (which expires in 2038)

...and have exactly the same...(plus we get all the sites the user went to as a bonus)

What prevents your company/government to do the same?
Or ... publish your proxy in open proxy list / become a TOR node

Collecting your search terms



I run a super secret project called **Sookah**.

I don't ever want people to know about it.

When someone search for the word Sookah I want to know it leaked to someone

I don't want them to find out that I know

I register an Adword...isn't Google wonderful?

Web Results 1 - 10 of about 449 for sookah. (0.14 seconds)

SoundClick song info: Sookah by Hadjee and the Wartones - Modern ...
 SoundClick band page for Hadjee and the Wartones: band bio and Traditional Arabic MP3 music downloads.
www.soundclick.com/bands/songInfo.cfm?bandID=471024&songID=4380477 - 16k - Supplemental Result - [Cached](#) - [Similar pages](#)

Lyrics for "Sookah" by Hadjee and the Wartones - SoundClick song info
 About "Sookah": the sun rose and so did she, with a renewed vigor and energy, determined to meet her daily destiny thru the trials and tribulations of the ...
www.soundclick.com/bands/Lyrics.cfm?BandID=471024&songid=4380477 - 13k - Supplemental Result - [Cached](#) - [Similar pages](#)

sookah's Favorites » Arab-Zone.com
 Arab-Zone.com. Meet Arab People, Make new friends... Blogs, Friends, Groups, Emails and alot more!
www.arab-zone.com/sookah/favorites/ - 16k - [Cached](#) - [Similar pages](#)

sookah's Profile » Arab-Zone.com

Sponsored Links

[Register now](#)
 Best priced domain and web hosting. Register your domain today!
www.register.com

Getting Started

Google Ad

Go AdWords

Campaign Account Snapshot

All Campaigns

+ Create a new campaign

| <input type="checkbox"/> | Campaign Name | Current Status | Current Budget [?] | Clicks | Impr. | CTR | Avg. CPC | Cost |
|--------------------------------|------------------------|----------------|---|--------|-------|-------|----------|--------------|
| <input type="checkbox"/> | Sookha | Active | R15.00 / day | 0 | 1 | 0.00% | - | R0.00 |
| Total - all 1 campaigns | | - | R15.00 / day active campaigns | 0 | 1 | 0.00% | - | R0.00 |

[Learn how your account settings affect your ad performance.](#)

Reporting is not real-time. Clicks and impressions received in the last 3 hours may not be included here. Time zone for all dates and times in data tables, reports, and billing: (GMT+02:00) Johannesburg. [Learn more.](#)

Starting page is: **Campaign Summary (this page).**
[Make Account Snapshot my starting page.](#)

A different thought



Your life story in no more than 5 pages
...A.k.a your resume'

Once you get someone's resume' you know all about the
person

You can search for it ...or...

You can get people to send it to you

Recruitment is easy:

Post a job ad and wait for people to send their life story

You can even specify which types of people....:)

*“Looking for nuclear scientist/engineer with experience in
Uranium enrichment and military background. Earn top
dollar, 401K plan, dental coverage, 25days leave. Flexi time.
Apply within...”*

Part II : Using the information



Hackers are not good at applying information.
They are devious – but not outright criminal.

I'll give it a shot though...

Using the information

Hit & run



- Spoof email from the FD to employees (& Bloomberg) stating the CEO has resigned / company is insolvent / sell your shares / etc.
- Register a site in the name of the holding company's director. Mirror a porn site. Populate. Spoof mail from a techie at the sister company to everyone about the “discovery”. Make it look like a CC that went wrong.
- Spoof SMS from the FD's mobile phone to a high profile investor about corruption in the company. Watch the share price drop, buy low, sell high.

But this is kid's stuff and is easy to spot. Timing however is everything...

Using the information



Back to company X

Let's assume we create a 'information footprint' of X using the framework.

After a couple of hours we know who are directors at company X and

- Their email addresses – their private ones too
- Their hobbies
- Their social network – clubs they belong to, their friends
- The location of their blogs

We also have a list of email addresses of their employees

So...

Using the information

Personal online identity theft



If these people don't have a strong online presence we will help them out:

- Create a Gmail (or pick any other widely used free email) address in their name
- Register them (and their friends) on LinkedIn
- Register them on MSN/Skype/Google Talk
- Register them on MySpace (or pick appropriate social network)
- Create a blog for them on Blogspot/WordPress
- Subscribe them to a couple of mailing lists
- Oh...and you might register a similar domain for them (like...phishing you know)

Can you automated this ? (even with CAPTCHAs?)

We've collected all the info to re-create them very life like

Using the information: identities are grown, not born



People don't appear from nowhere

We need to give their new identities credibility

If you Google someone...they magically exists

Blogs and posts are time stamped

- Unless we control the underlying infrastructure (although its more effort and a single point of failure)

Thus – we need to grow these identities because we can't manipulate the time line

- Arb postings to mailing lists/blogs
- Sign a “guestbook” (remember those?)

Basically get the identity out there on anything that's indexed – and will thus show up later.

It can take months...perhaps years..if you do it right

If we investigate it – it must look real.

Using the data

Starting a campaign



- Now that I am you and your board I need a mission
- It actually boils down to plain old marketing
- People are flock animals
- So first create a virtual flock of people.
 - Grow many more identities (we know the players)
 - Manipulate counters (its SO 90s)
 - Manipulate Votes / polls
 - Generate automated comments / forum
 - ...and others will follow

Seems like a lot of hard work, but
...watch this space...

Coming soon to blog near you: automated comments



/code/autosent roeloftemmingh\$ perl generate.pl data-disagree/

- I don't understand this. The ideas you spoke about are honestly bullshit. I have never seen such a load of bullshit.
- Our company basically do not figure your statement. The fundamentals here are wrong! I have not experienced this type of situation! Do you believe this nonsense? This is totally fabricated.
- I have surely never seen this type of situation. Get a life. You honestly believe in this nonsense? This is FUCKING appalling.
- We do not get this thinking.
- This idea is unbelievably bad! You consider this waste? This is nonsense!
- FUCK THAT!
- This posting is bad! The thinking here is full of shit.
- I have never ever seen any of the things listed here. Dude, you should be spending time on other stuff!
- This is fucking ridiculous. Do you consider this bull? C'mon - think about it.
- Your company ought to be researching more interesting things.
- We utterly do not figure this article. This argument is rubbish.

CAPTCHAS??



Go watch “Wag the dog” again...

...but think of the Internet of today.

Thus...in conclusion

The mouse is mightier than the pen



- Security experts tend to focus on technology itself, ignoring the application and surroundings of it's use.
- The web 2.0 contains great tech (?secure?) but little is known about the security implications when the tech is actually used.
- Real criminals don't write buffer overflows – they follow the route of least resistance.
- Mainstream criminals tend to lag behind. We knew about phishing attacks back in 95.
- What will be on their minds in 2010?
- I am guessing it would be something close to this...