Investigating Individuals & Organizations Using Open Source Intelligence

A NEW TRAIN OF THOUGHT
PATERVA

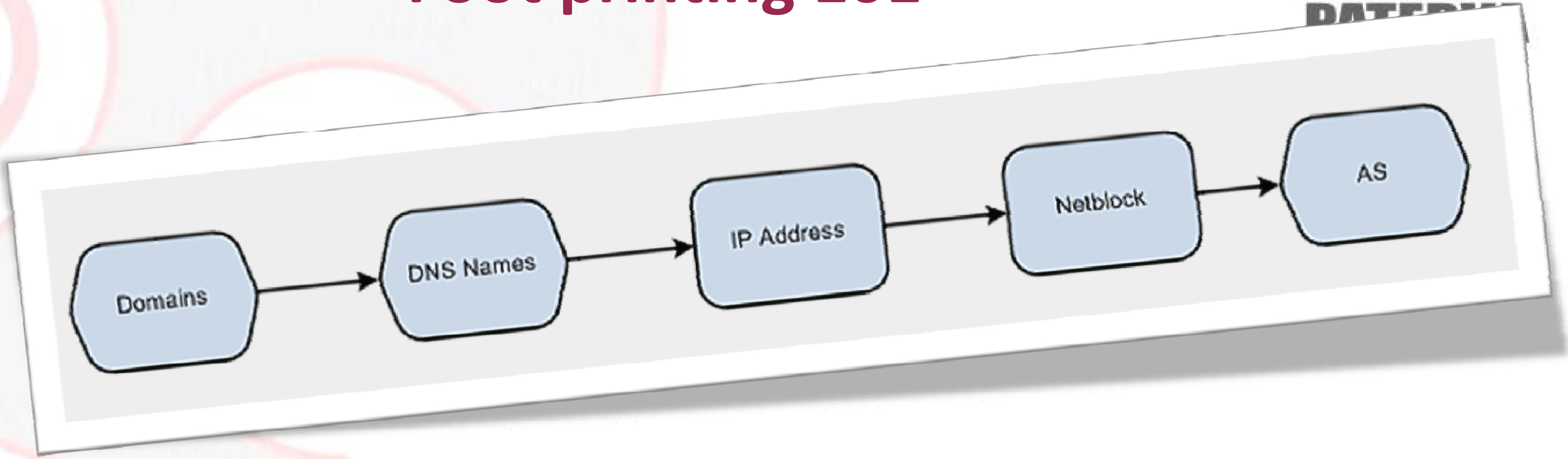
Introduction

Who are we?

Roelof Temmingh
Paterva (http://www.paterva.com)
roelof@paterva.com

Chris Böhme
PinkMatter (http://www.pinkmatter.com)
chris@pinkmatter.com

Foot printing 101



Four conversions or transforms:

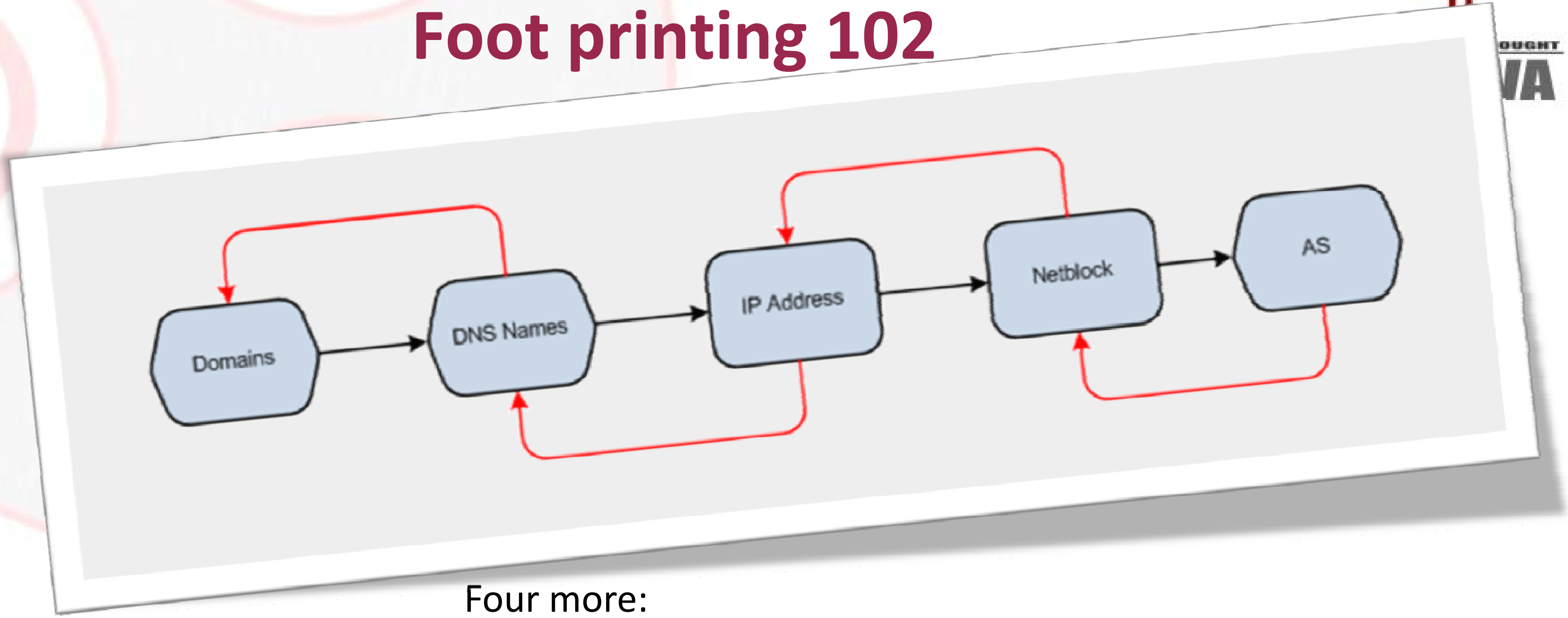
Domain to DNS Name
MX/NS/Zone transfer/Brute

DNS Name to IP address
Just resolve

IP address to Netblock
Whois

Netblock to AS
core routers, web

Foot printing 102



Four more:

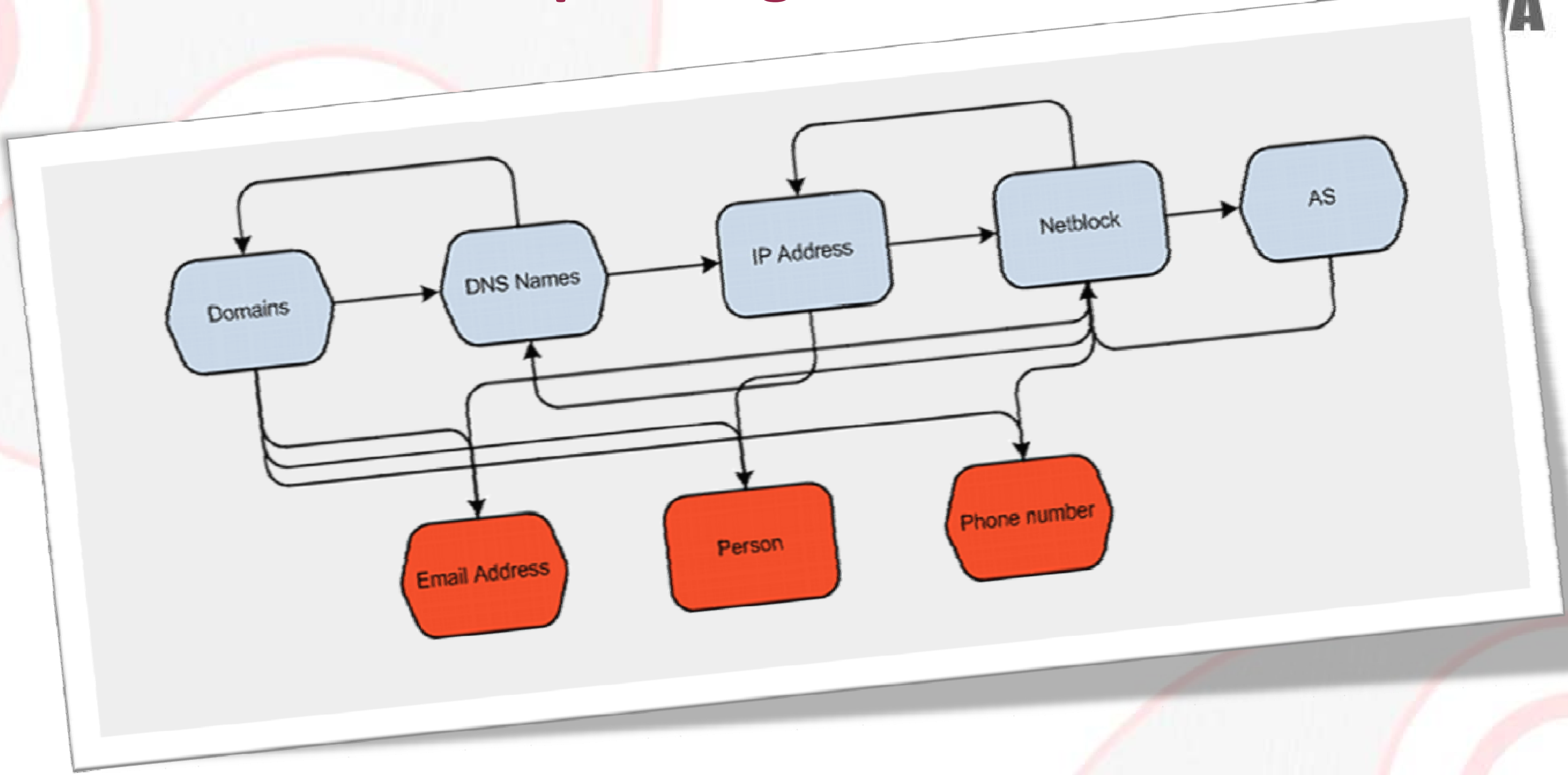
AS to Netblock
Robtex / Cymru

Netblock to IP address(es)
Just expand

IP address to DNS name
Reverse DNS, shared virtual hosts etc.

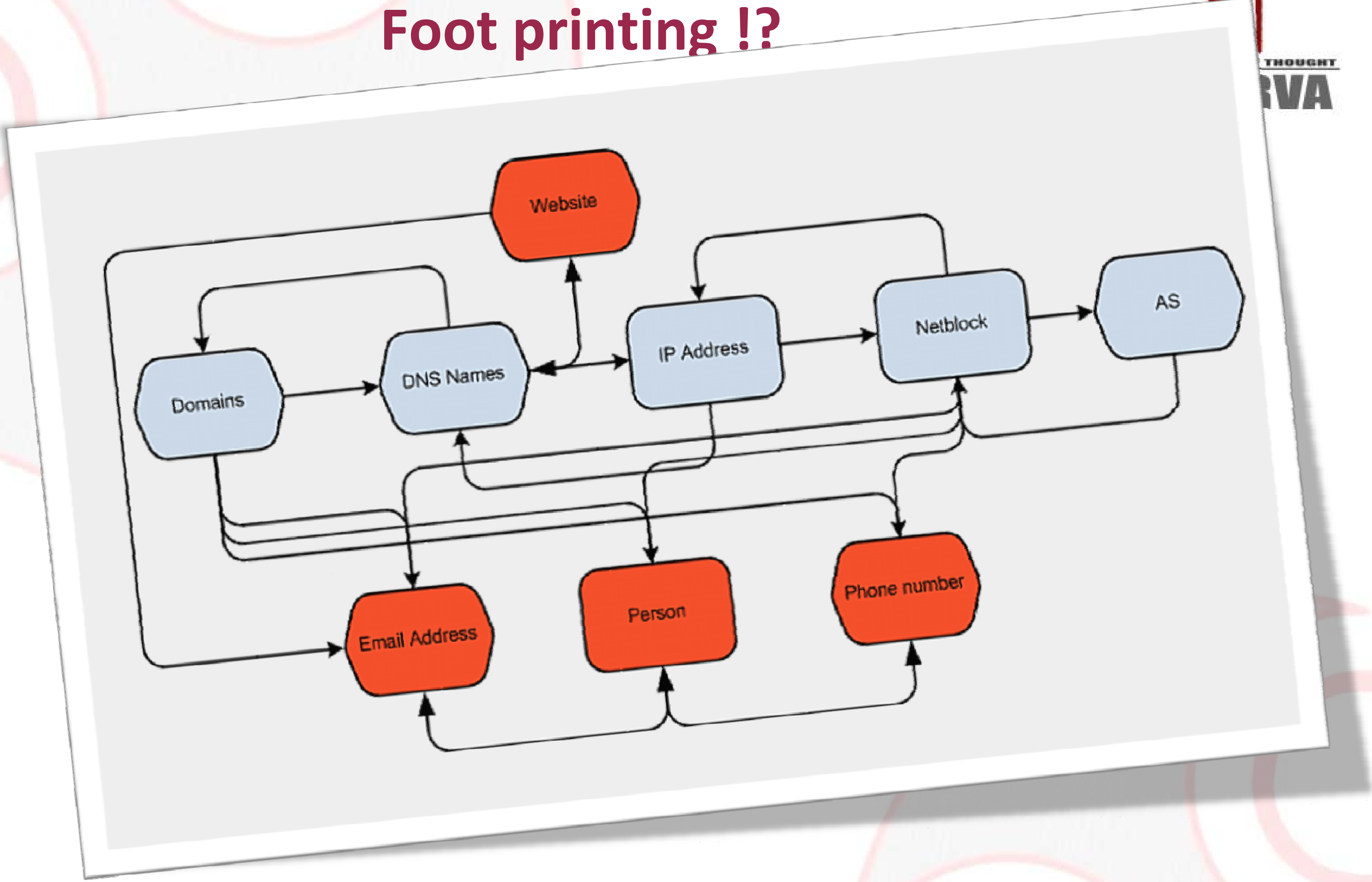
DNS Name to Domain
Simple..?

Foot printing 201



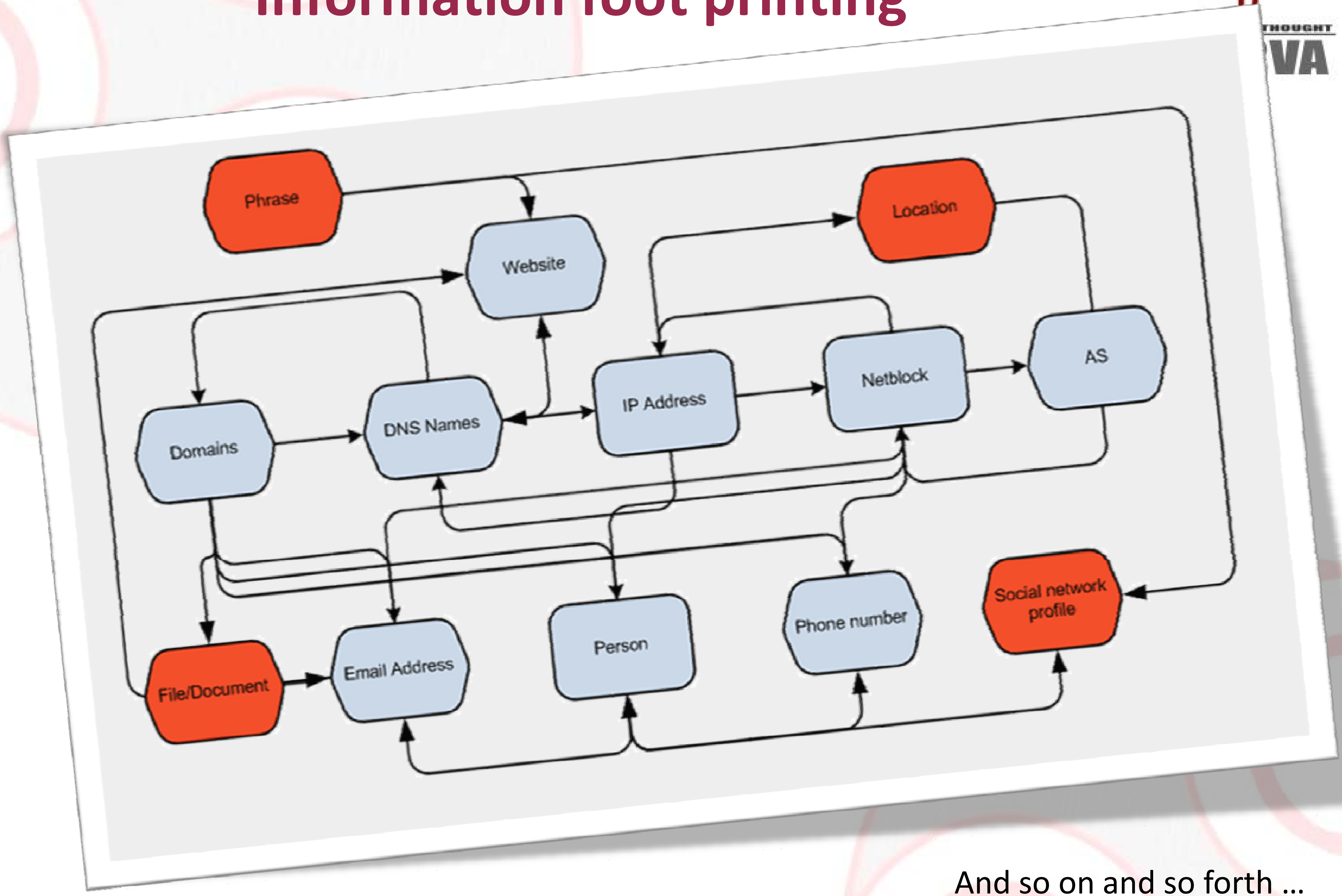
Six more transforms….via Whois

Foot printing !?



And six more…using Search Engines, PGP etc.

Information foot printing



And so on and so forth …

A container of tricks

Almost every arrow represents:

- Some kind of trick, cute algorithm
- Something to keep in mind
- A bookmark or a friend

You end up with loads of scripts / methods / apps
The need for a box of tricks became evident

The need for a GUI

Also :

- If A -> B -> C and X -> Y -> C, then X =~ C
- A -> B -> C -> A
- Keeping track of where we’ve been

Seeing this without a graph representation is almost impossible

Demo of Maltego

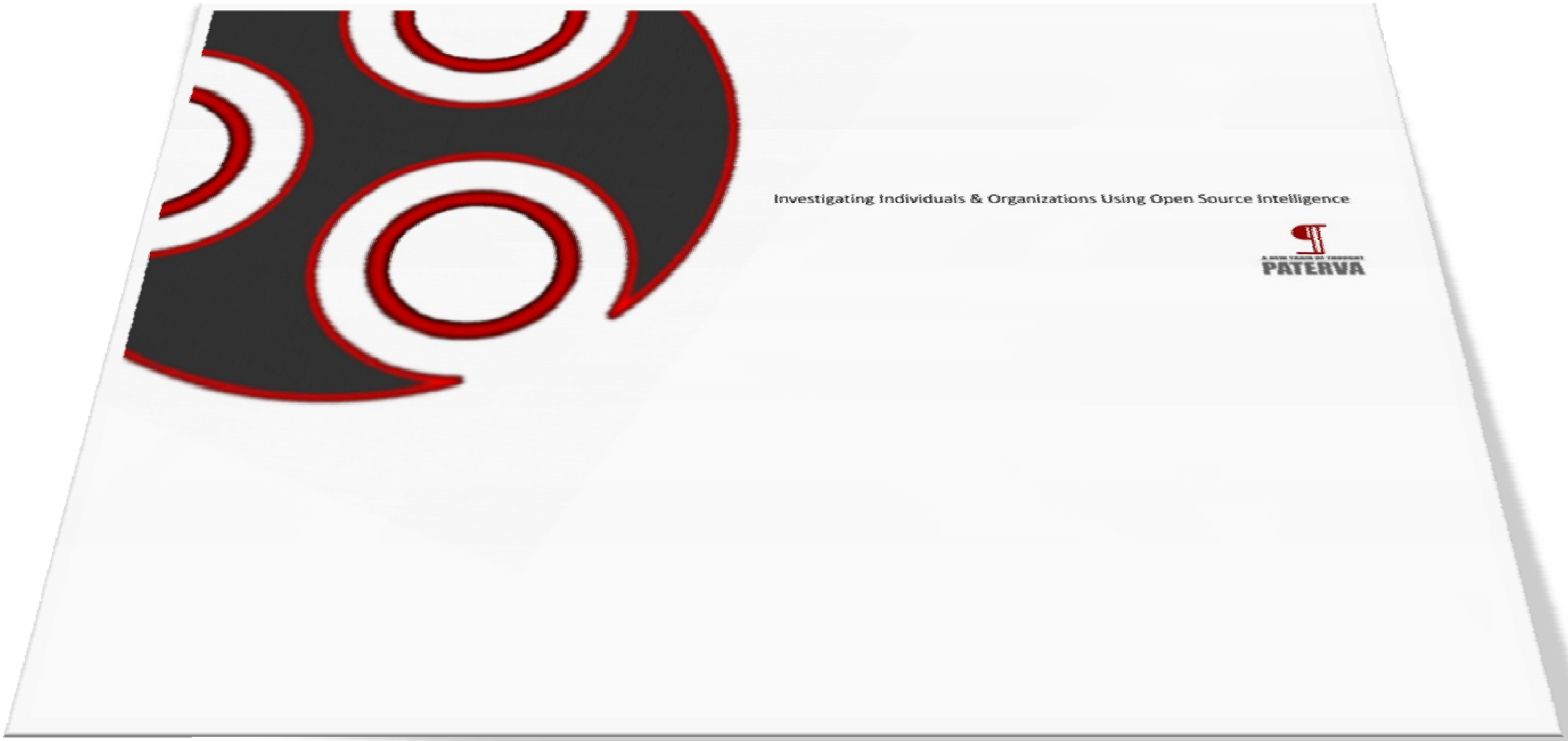
What is information really?

- Information or just data?
- Too much or too little?
- What are we really looking for?
- Humans are
  - Great at recognizing patterns
  - Lousy at processing data
- The challenge: The human friendly middle way

A NEW TRAIN OF THOUGHT
PATERVA

Demo of Maltego: Information Zen

See also Managing Complexity



Investigating Individuals & Organizations Using Open Source Intelligence

PATERVA

Walkthrough of a Transform

1. Get entity from user
   Person – Roelof Temmingh
2. Get question on entity from user
   convert to email address
3. Expand question and add confidence levels
   Roelof Temmingh, Rtemmingh, TemminghR etc.
4. Ask the question to your data source
   search for it on search engine*
5. Get the answers
6. Parse the answers for output entities
   parse for email addresses [at/_at_/remove]
7. Process the parsed entities
   confidence, frequency, correlation etc
8. Show top N processed entities

In many cases it’s a 1:1 relationship, so confidence levels etc do not apply
*more later…

Challenges

Operational

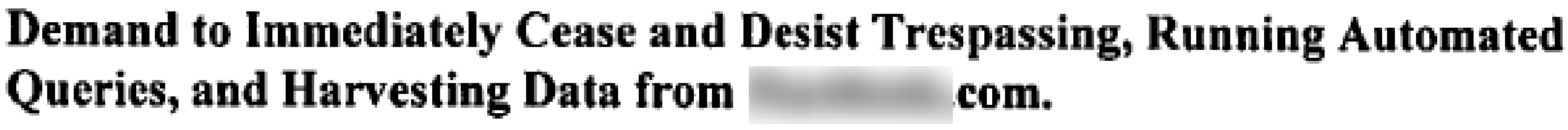
- Not everyone gives their details out on the Internet (yet)
- Information on the Internet is not clean – there are no standards
- It’s really hard/impossible to give context to the information

Technical

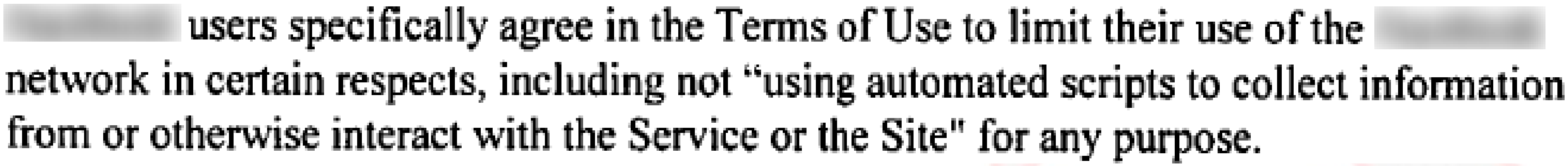
- Some entity types are really hard to parse effectively
  - Person
  - Phone number
- Speed

Legal challenges

- Information - the currency of many sites
- They will give you the info, but on their terms:
  - see their ads
  - submit your personal details
- Even APIs restricted to
  - personal use
  - limited queries
- Automated collecting/scraping is prohibited



Demand to Immediately Cease and Desist Trespassing, Running Automated Queries, and Harvesting Data from [redacted].com.



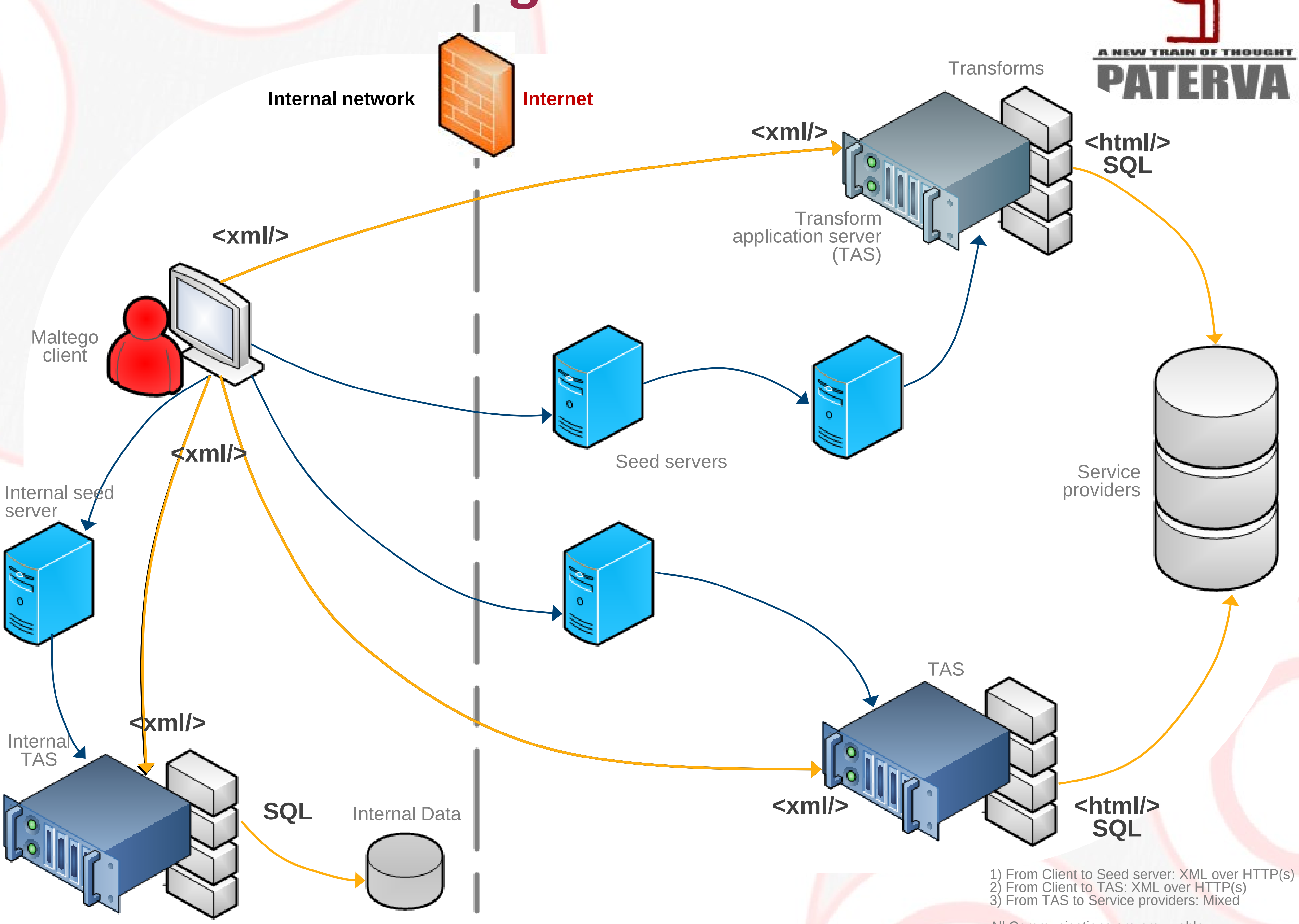
[redacted] users specifically agree in the Terms of Use to limit their use of the [redacted] network in certain respects, including not “using automated scripts to collect information from or otherwise interact with the Service or the Site" for any purpose.

Distributed transforms

- No more transforms shipped with Maltego app
- Sea of public transform servers

- Other benefits
  - Building a community of transform writers
  - Anyone, any-how
  - Simple interop (XML, HTTP)
  - High scalability (bandwith, caching)

Maltego architecture



1) From Client to Seed server: XML over HTTP(s)
2) From Client to TAS: XML over HTTP(s)
3) From TAS to Service providers: Mixed

All Communications are proxy-able

But…it didn’t go away…

- Moved from Google to Yahoo API
- Cut 32 transforms from public TAS
- !Social networks transforms → Rapleaf, Spock
    - Not real time, not comprehensive

- Yahoo API key limits

Legal challenges II

"Building a tool around scraping Google sounds like a good way to make broken software“
- someone at Google

Will you revive the API? - No
Can I buy (pay per click) access somehow? - No
Will it help if I show your ads? - No
Will rate limiting my requests be helpful? – No

"It's the users who lose out when Web companies decide to crack down on popular scrapers"
- Reid Hoffman, CEO of LinkedIn

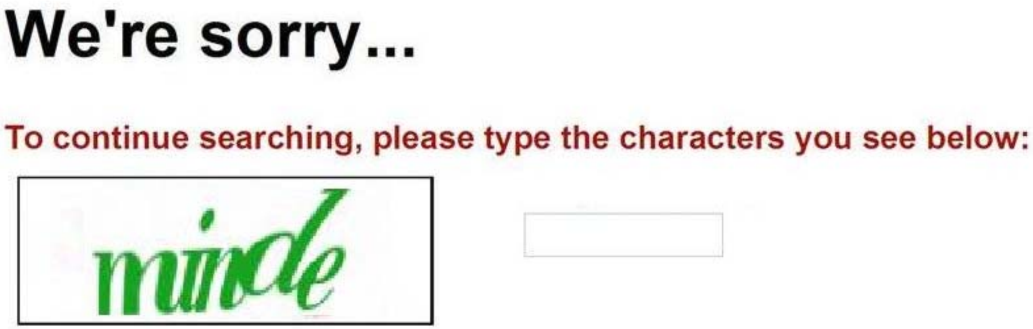
Bottom line – Any amount of technical cleverness, thinking and experimenting just won’t help. ‘You can’t do it because I said so’.

The Bakery

“**you can’t do it because I said so**”

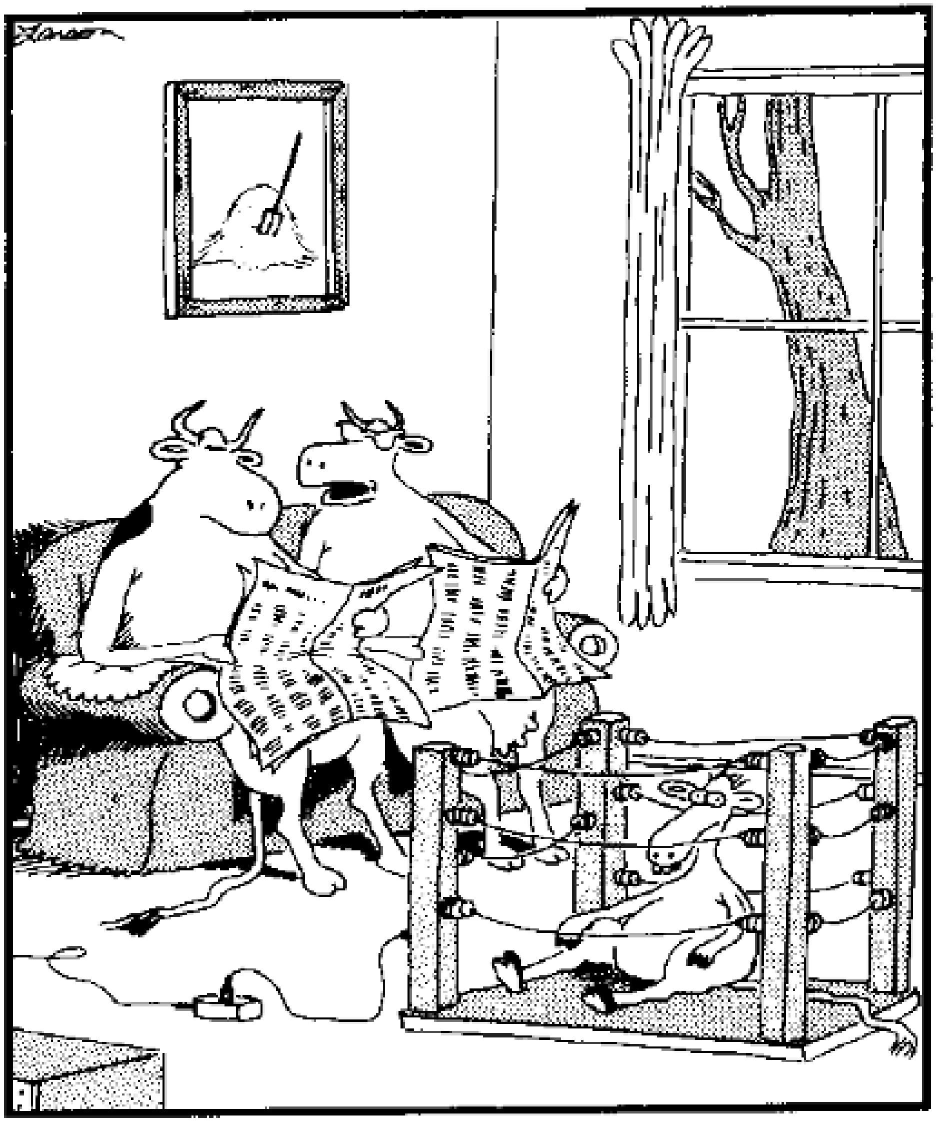
Usually not a good idea to say that to security people…

…but



We're sorry...

To continue searching, please type the characters you see below:



"Ben — what d'ya say we turn the power off for a while and let the little guy roam around?"

Investigating Individuals & Organizations Using Open Source Intelligence

A NEW TRAIN OF THOUGHT

PATERVA

More human than human

If you can convince an algorithm that you are human, can you convince a human that you are human?

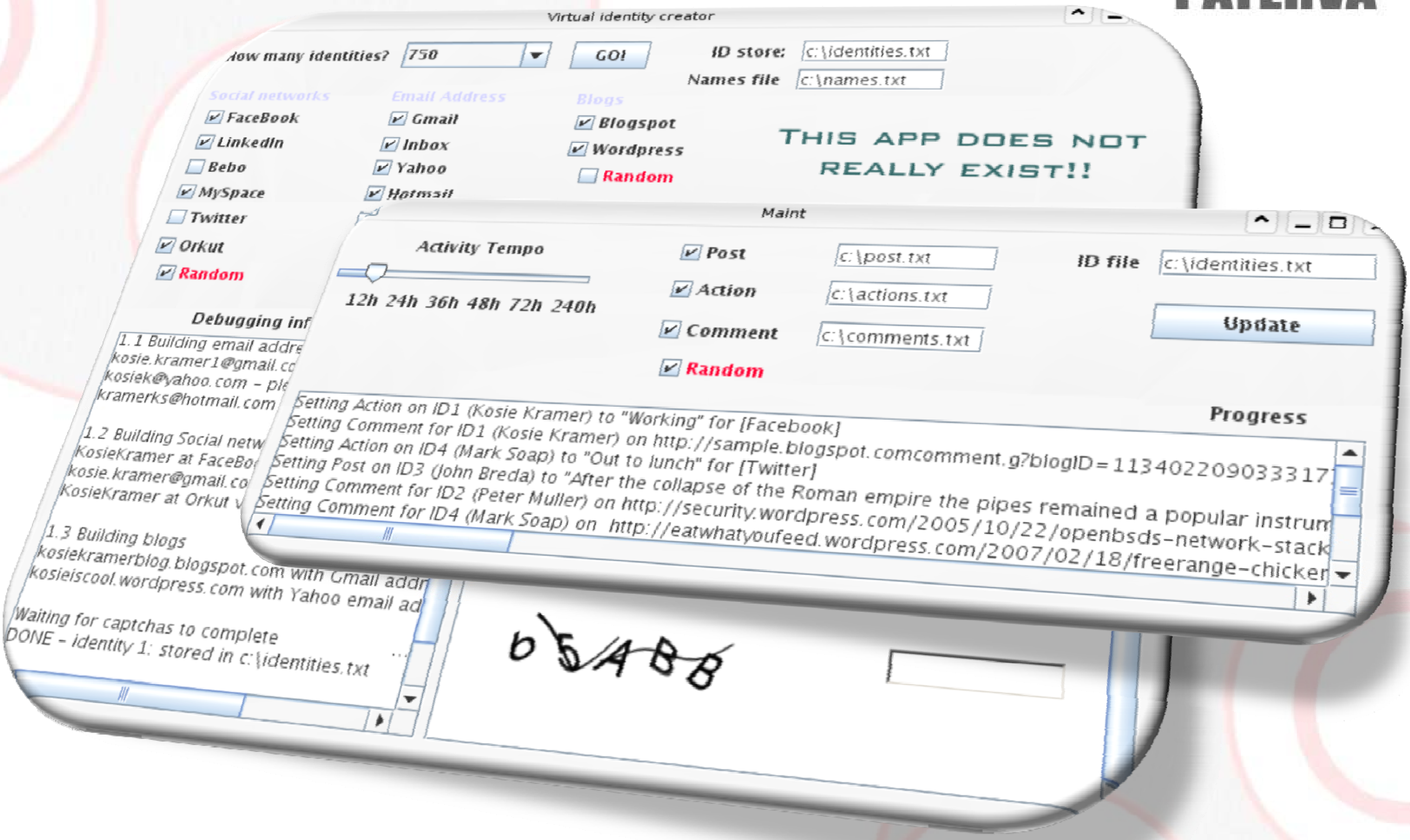
Consider CyberLover, a Russian chat bot – communicating with users over IRC – thus, in real time and interactive:

Among CyberLover's creepy features is its ability to offer a range of different profiles from "romantic lover" to "sexual predator." It can also lead victims to a "personal" Web site, which could be used to deliver malware, PC Tools said.

We've collected all this nice info with Maltego, what can we do with it?

Making imaginary virtual friends

A NEW TRAIN OF THOUGHT
PATERVA



Virtual identity creator

How many identities? 750 ▾ GO!

ID store: c:\identities.txt

Names file c:\names.txt

Social networks
- ☑ FaceBook
- ☑ LinkedIn
- □ Bebo
- ☑ MySpace
- □ Twitter
- ☑ Orkut
- ☑ Random

Email Address
- ☑ Gmail
- ☑ Inbox
- ☑ Yahoo
- ☑ Hotmail

Blogs
- ☑ Blogspot
- ☑ Wordpress
- □ Random

THIS APP DOES NOT REALLY EXIST!!

Debugging inf

```
1.1 Building email addre
kosie.kramer1@gmail.co
kosiek@yahoo.com - ple
kramerks@hotmail.com

1.2 Building Social netw
KosieKramer at FaceBoo
kosie.kramer@gmail.co
KosieKramer at Orkut v

1.3 Building blogs
kosiekramerblog.blogspot.com with Gmail addr
kosieiscool.wordpress.com with Yahoo email ad

Waiting for captchas to complete
DONE - identity 1: stored in c:\identities.txt
```

Maint

Activity Tempo

12h 24h 36h 48h 72h 240h

- ☑ Post c:\post.txt
- ☑ Action c:\actions.txt
- ☑ Comment c:\comments.txt
- ☑ Random

ID file c:\identities.txt

Update

Progress

```
Setting Action on ID1 (Kosie Kramer) to "Working" for [Facebook]
Setting Comment for ID1 (Kosie Kramer) on http://sample.blogspot.comcomment.g?blogID=113402209033317
Setting Action on ID4 (Mark Soap) to "Out to lunch" for [Twitter]
Setting Post on ID3 (John Breda) to "After the collapse of the Roman empire the pipes remained a popular instrum
Setting Comment for ID2 (Peter Muller) on http://security.wordpress.com/2005/10/22/openbsds-network-stack
Setting Comment for ID4 (Mark Soap) on  http://eatwhatyoufeed.wordpress.com/2007/02/18/freerange-chicken
```

Exploiting

Quantifiable results:

- Counters
- Ratings (it’s just SO web 2.0)
- # of users (what if 75% of your user are bots)

This is really click fraud if you think about it

But also more fuzzy results (hey it worked for interactive sessions!):

- Positive/negative comments on an article/blog
- Opinions in an article, blog posting
- Tags
- IM status lines

The players here have a vastly different skill set than what we have. Web 2.0 (urghhh) is the vulnerability and the content is the payload.

Peer pressure

So what can we do with it?

- Manipulate ratings of anything
- Sway public opinion
- Influence political polls
- Alter stock prices – directly or indirectly
- Perform social denial of service

Keep in mind that people are flock animals – you just need to be the initial catalyst and get critical mass

Why are we at BlackHat?

Making the application do something it shouldn’t do
vs.
Using the effects of using the application for interesting purposes

Some applications just shouldn’t have been built:

- Bank giving their users free email
- Sending proof of payment via email
- School friends

The (perfect normal use) of the application leads to vulnerability

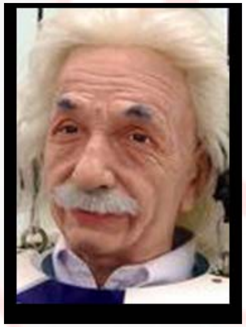
Speed of technical assessment
vs.
Speed of assessing the effects

Conclusion

Have we *hacked* anything actually ?

You know you screwed up when you want to go back and try to undo/regulate/un-invent what you've done:

- The atom bomb
- Chemical warfare
- Cigarettes
- Facebook



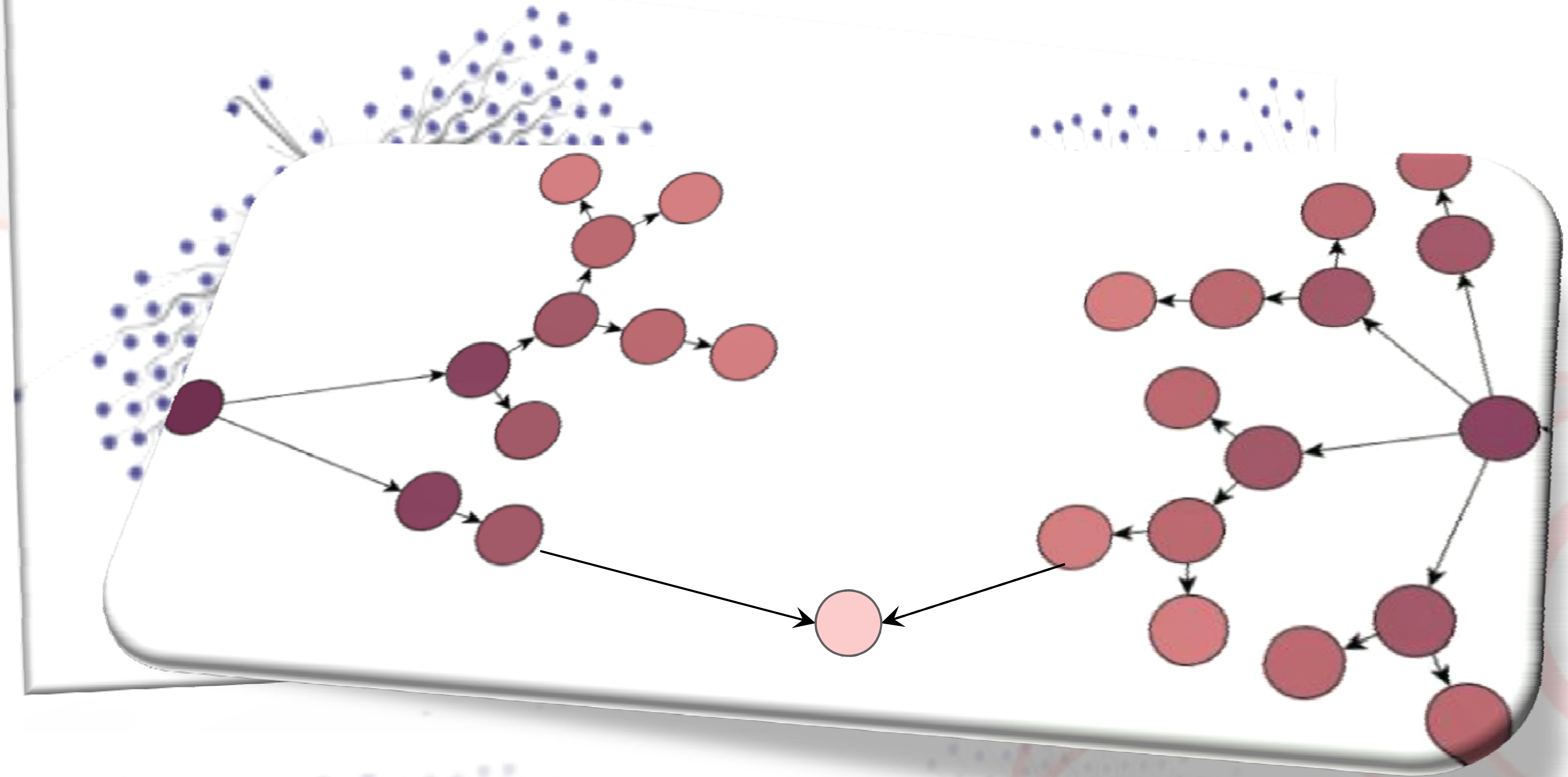
Appendix

Investigating Individuals & Organizations Using Open Source Intelligence

PATERVA

Managing Complexity

- Data becomes relevant when links exist
- Hidden nodes, hidden links



Back