# Social Networking Sites:
# Data Mining and Investigative Techniques

Steve Patton, CISSP

The opinions expressed herein are solely those of the author.

# Common Sense

Avoid child porn by not downloading graphics.

Pay attention to Terms of Service requirements.

Do not bypass site controls.  This is not cracking.

Be a good netizen – do not overload your target.

# Social Networking Sites

Based on site structure:

| | |
|---|---|
| MySpace | easy to mine, but messy HTML |
| Friendster | easy to mine |
| Bebo | easy to mine |
| Xuqa | moderate |
| Facebook | difficult: must be friend |

# Social Networking Sites

Based on Google results:

| | |
|---|---|
| MySpace | 93Million |
| Friendster | 1Million |
| Bebo | 7.5Million |
| Xuqa | -- |
| Facebook | 300K |
| Xanga | 9M |

# Information available on SNS

Friends, interests, contact information, historical information, pictures and more...

Frequently the information is unstructured, and random.  It can also be outdated or inaccurate.

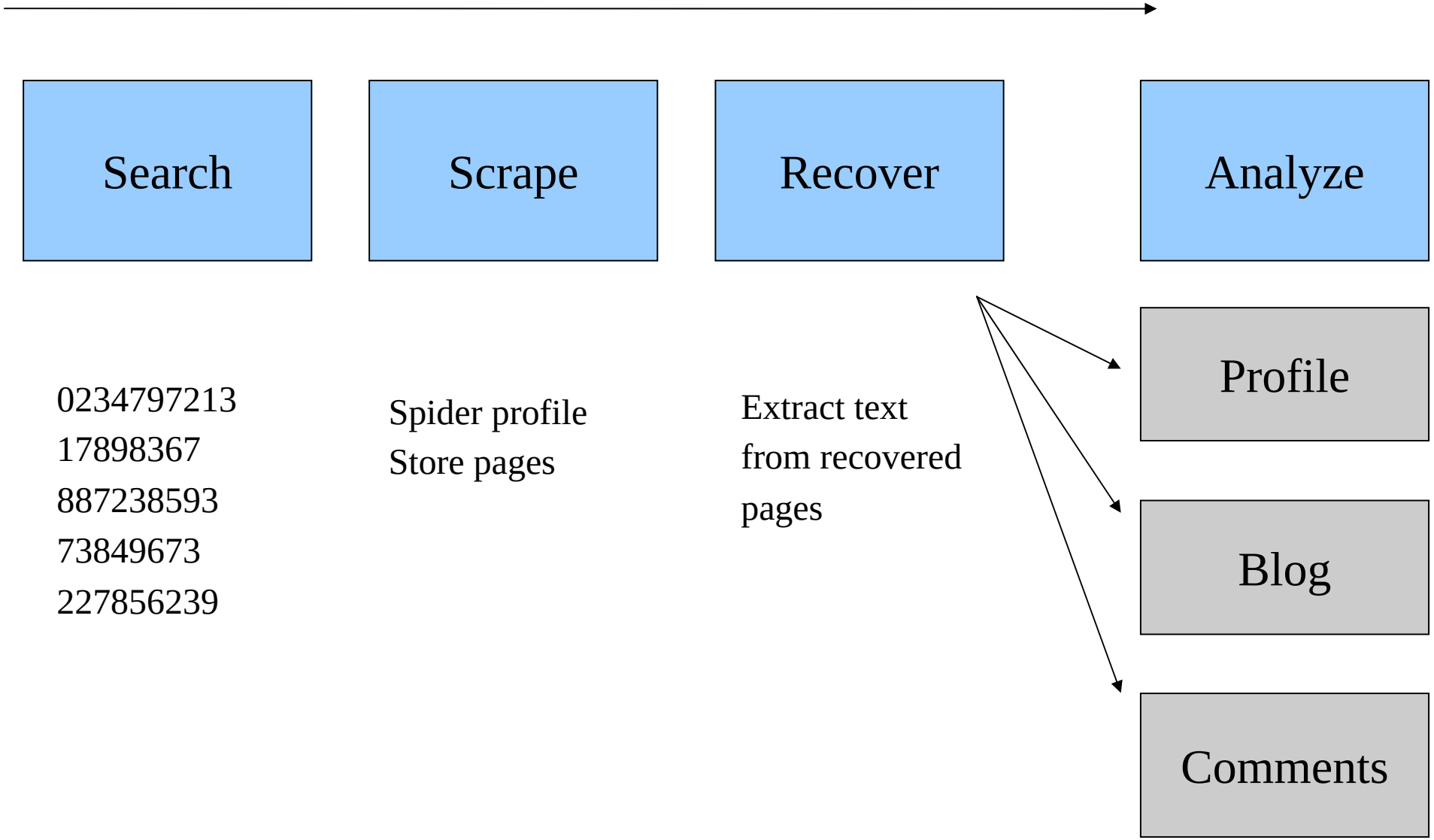
# Data Mining on SNS

Site structure is key:

Profiles are generally linked by an account # or ID

Profile contains
- profile page
- blog page(s)
- picture page(s)
- comment page(s)

# Data Mining on SNS



# Factoids

2 runs against identical search criteria:

September 2006

- 774 profiles
- 750M recovered
- 140K comments

April 2007

- 698 profiles
- 953M recovered
- 145K comments

# Investigating

Remember that profiles are dynamic.  Many users change their profile daily.  Print or save copies as needed.

The web is highly linked.  Follow the links to get a better picture of what you are looking for.

There are many millions of pages.  Hone your search or you will waste your time.

# Searching

Check search engine caches!

The cached page may hold gold while the current page only has coal...

Different engines may have different results, though it is TOUGH to compete with Google!

# Signing In

Many sites restrict content to members only.  To see more content, you have to register and sign in.

Possible drawbacks include:

- Maintaining an ID and password
- Loss of anonymity
- Tipping off your target
- Ethical issues around terms of service
- Evidentiary issues around investigatory methods

# Viewing Tip 1 (Easy)

Frequently profiles are hard to read due to backgrounds and color schemes.

- Drag your mouse across comment text.
- Select the text by holding the left mouse button as you drag.
- The selected text will change colors perhaps becoming readable.

# Viewing Tip 2 (Medium)

(In case the previous tip failed...)

- (Firefox/Mozilla) From the View menu, choose View->Page Style->No Style
- (Internet Explorer) From the Tools menu, choose Tools->Internet Options->General and then select fonts/colors that improve page readability.

# Viewing Tip 3 (Hard)

- Right click while hovering over the background
- Select “view source”
- Save the resulting file
- Edit the file, removing from <style> to </style>
- Save and view the resulting file

# Pictures

Pictures are a valuable part of a profile.  Find more by following links and observing URLs in the HTML.

Few photo web sites provide directory listings, but if you look through the HTML, you'll find some that do!

# Permissions

“Private Profiles” or Sign-in pages (logon screens) are generally dead ends indicating the remaining content is protected.  Don’t tamper with the lock!

# Some photo hosting sites

(Your mileage may vary)
imageshack.com – no albums
tinypic.com – no albums
photobucket.com – depends on the album owner
916online.com – wide open

You'll have to experiment to determine current status of these and others.