Every Breath You Take
::data gathering, persistence, and what it all means::
One of the overriding themes of this talk is that once a piece of data has been captured, it’s forever retrievable. In the spirit of that principle, these made-for-CD slides have been watered down a little bit, in order to prevent any unintentional long-lived copyright violations, overly offensive topics or other goof ups.

Dynamic, up-to-date slides can be pulled from: http://dl.jimio.com/dc0f/ebyt.pdf
Abstract

• How much data do you generate in the process of living an ordinary day? This talk covers various ways to gather, persist and analyze the data stream that is your life. We’ll cover a few of the approaches that are available today, some easy code you can whip up to persist anything you please, and what to expect from the community and businesses moving forward. Privacy/security impact is sure to be huge, so hold on to your hats, and start tracking and logging everything! Somebody else may be doing it for you already.
TODO: Data *

• Definition
• Gathering
• Persistence
• Mirroring
• Analysis
• Forecasting / Targeting
Allow myself..

- Jim O’Leary, aka jimio
  - In the security game since about 2000
  - CompSci & Psych background
  - Mostly whitehat / corporate stiff
  - Enjoys: design, coding, breaking code, long walks

- Contents of this talk do not reflect the views of my employer, friends, countrymen, etc.
Data: Definition

STUFF

TIME
LOCATION
HEIGHT
TEMPERATURE
BMI
WEIGHT
PURSE
RADIATION
BAG
DNA
DATE
REALITY
MANIPULATED LEVEL
STRENGTH
RELIGION
FINGERPRINT
PURCHASE HISTORY
EYESIGHT
Data: Classification

PRIVATE
INHERITED
INTERESTING
SPOOFABLE
COMBINABLE
PUBLIC
PERSONALLY IDENTIFIABLE
ENVIRONMENTAL
UNINTERESTING
Data: Awareness

• Information is your friend!
• Knowledge is power!
• The Vegas Hustle
  – Try to find a clock anywhere near a game table
  – Abstract tokenization of your cold, hard cash
  – Endless free drinks to modify perception
  – Chummy pit bosses, scantily-clad waitresses
  – Overwhelm and distract, only to take your money
## Data: Gathering

### Acquisition Methods / Data Types

<table>
<thead>
<tr>
<th>Method</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Push</td>
<td>email</td>
</tr>
<tr>
<td>Pull</td>
<td>RSS</td>
</tr>
<tr>
<td>Voluntary</td>
<td>Twitter</td>
</tr>
<tr>
<td>Involuntary</td>
<td>ISP snooping</td>
</tr>
<tr>
<td>Manual</td>
<td>Pen and paper</td>
</tr>
<tr>
<td>Self-regulating</td>
<td>Whitelisted / targeted</td>
</tr>
<tr>
<td>Self aware?</td>
<td>Intelligent sync, data resolution</td>
</tr>
<tr>
<td>Inadvertent</td>
<td>Whoops!</td>
</tr>
</tbody>
</table>
Data centres: Every spy should have one

from New Scientist Technology Blog by Tom

If you used the instant-messaging system MSN messenger in June 2006, all your chats were being collected and being passed (anonymised, of course) to researchers at Microsoft Research, a paper in the physics pre-print server arXiv reveals.

But their results aren't headline grabbing - they show people are more likely to chat with others in the same geographical location, age group and of the same sex. As the arXiv blogger points out, more interesting is the fact the researchers struggled with the size of their dataset:

"The sheer size of the data limits the kinds of analyses one can perform," they write. The 4.5 terabytes of chat logs took 12 hours to copy onto an eight-processor server for processing.

The US security services have made moves to assemble databases of online communication that will surely dwarf that - watching over phone calls, social networking sites and emails.

Extracting useful information - not just generalities like the study mentioned above - is going to require massive amounts of storage and processing power. The spying game just isn't the same any more. How long before James Bond is seen acting on a tip-off from network theorists with ranked armies of servers at their backs?

Tom Simonite, online technology reporter
Data: Gathering - Fun

- We are in Vegas, after all..

(quick contest, w/ prizes : details announced mid-talk)
Data: Harvesting

• Massive server farm would be nice and all..
  – Expensive
  – Time consuming

• ..so have somebody else do it for you
  – GOOG, MSFT, YHOO, INSP..
  – Just speak HTTP/HTML, and you’re good to go
  – Quick demo
Data: Persistence

• A moment like this
  – Digital photography’s influence
  – Huge increase in the # of data-feeding nodes

• What’s been done can’t be undone, son
  – Sure, you can remove the “tag” in the photo
  – Sure, you can delete those pictures from your camera
  – Once it’s been captured, it’s out of your control forever
Social Networking

• Facebook
  – Digital representation of my social life
  – News feed to passively monitor your friends
    • Who’s doing what, where?
    • Image tagging / Facebook paparazzi
    • The “in a relationship” struggle

• Myspace
  – Friends, Bands, Bulletins, Events
  – Glittery roses, pouty lips, XSS worms
Security Through Obscurity

Privacy Through Obscurity
Data: Privacy

• Don’t want to get caught?
  – Don’t do it
  – Do it, lie about it
  – Do it, try to delete all the evidence
  – Do it, tell all your friends in person
  – Do it, tell your friends online
  – Do it, take pictures
  – Do it, post pictures online
  – Do it, lie about it, change mind, issue press release
The Sex Tape Era

[more examples]
Data: Elusiveness

• Need to send something?
  – Do it on your own terms
  – Encrypt!
  – Minimize server hops/exposure

• HTTPS, just do it.
Data: Mirroring

• The “moment in time” principle
  – *Nowness* tied to data value
  – Presence seems ok
  – Presence history is just creepy
    • That’s why you do it yourself..

• Persononories
  – Digital representations of self
  – Skip the resume, hit up Google
Data: Analysis

• Now that we’ve got it, what do we do?
  – Identify patterns
  – Find trends
  – Simulate and model situations
  – Profit.
Analysis: Some Fun Demos

• Can a brother get a _____?
  – Cliché / snowclone iteration finder

• For Unlawful Carnal Knowledge
  – Article Parser, Pattern Recognition

• Inadvertent Haiku Finder
  – Article Parser, Natural Language Processor, Pattern Recognition
Data: Analysis

• Passive, non-invasive time travel and modeling
  – Way out there.. Matrix-type stuff
  – Given enough data to represent the physical world
Data: Forecasting / Targeting

• Marketing’s wet dream
• Actually, a ton of consumer benefit
  – Providing it’s done correctly
  – Also a ton of embarrassment potential

• Social network habits and dating inevitability
Data: Misc

• Few loose ends here..
EOF