Introducing Traffic Analysis

Attacks, Defences and Public Policy Issues...

George Danezis

K.U. Leuven, ESAT/COSIC,
Kasteelpark Arenberg 10,
B-3001 Leuven-Heverlee, Belgium.
George.Danezis@esat.kuleuven.be

29 December 2006, 23C3

Outline

A talk about 'Traffic Analysis'.

- Military techniques and defences.
- Attacking Internet Security.
- Extracting location & high level intelligence.
- Defences and attacking hardened systems.
- Future directions.

What is Traffic Analysis

Making use of (merely) the traffic data of a communication to extract information. As opposed to 'interception' or 'cryptanalysis'.
What are traffic data?

- Identities or call signs of communicating parties.
- Time, duration or length of transmissions.
- Location of emitter or receiver.
- No content – it may be encrypted.

A controversial starting point.

- Diffie & Landau statement – 'Privacy on the line' on the politics of encryption.
- "Traffic analysis, not cryptanalysis, is the backbone of communications intelligence."
- Could this become true on the Internet? Is it already?

Two views on military Sigint and information warfare (I)

"These non-textual techniques can establish

- targets' locations,
- order-of-battle and
- movement.

Even when messages are *not being deciphered*, traffic analysis of the target's C3I system and its patterns of behavior provides indications of his

- intentions and
- states of mind"

– Herman (JIC Chair)

Key points: no need for cryptanalysis, hierachy of sources according to availability / quality.

RF Direction Finding $\rightarrow$ Traffic analysis $\rightarrow$ cryptanalysis.

Two views on military Sigint and information warfare (II)

W.Diffie on Information warfare: 'Is it just a bomb in the computer room? – No!'
Understanding the command and control cycle:

- Intelligence → command → execution → intelligence . . .
- Contrasting examples: air operations in WW2 and Iraq ('91).
- The smaller the cycle the better (faster reaction, adaptability, efficiency, . . .)
- Attacks try to exploit the need to communicate fast and efficiently.
- Defences are costly and widen the cycle!!!

Relevance: Traffic Analysis resistance is *very expensive*. Unlike cryptology *defences are fragile*.
Civilian examples: Competitor Intelligence in taxis fleet management or Police communications.

The military world – concrete attacks

- WWI: Tapping the earth returns of telegraph lines.
- Naval and air operations: observing wireless communications. Radio silence. Morse 'hand'.
- Reconstruction of network structure of the German Air Force radio in 1941 by the British.
- Identification of radio equipment (also possible for GSMs).

Why is traffic analysis so valuable?

- It provides lower quality information compared with cryptanalysis, but it is both easier and cheaper to extract and process.
- Often used to perform 'target selection'. 'Economics of surveillance' (GCHQ and the Internet...)

Note the importance of *jamming* communications!

The military world – defences

Low probability of intercept and *Low probability of direction finding* communications.

- Principle: make the adversary spend time or energy (power) to detect and jam.
- Frequency hopping – modulate frequencies according to keys. Difficult to jam!
- Spread spectrum – transform signal to high band low power. Difficult to detect (under the noise floor).
- Burst communications – meteor scatter.
- Most important: clear security policies, doctrine, training, liveware. . .

Technologies used for civilian purposes: GSM (hopping), ADSL (SS), Cheap but reliable comms (meteor).
Reference: Ross Anderson, *Security Engineering*.

And then came (not just) the Internet. . .

Different environment – (not so) different players.

- Not so hostile – commercial use, personal use, government, (critical infrastructure?), (military?)
- A confederation of networks – different jurisdictions and security domains.
- Different transport technologies: cable, wireless, satelite, ATM, ethernet, . . .
- Common routing protocols – they expose traffic data.
- New technologies: wireless, overlay networks, convergence with telephone – more opportunities for collecting traffic data.
- Threats rapidly escalate – attack scripts!

Use of encryption – NG Telephony and Internet: traffic analysis only option (but also facilitated).
How to attack established security technologies? (Without making use of cryptanalysis or content)

Can the Secure SHell (SSH) protect your privacy?

SSH is used for secure remote login and file transfer. All data is encrypted and authenticated. What information can we extract about a password typed in a protected session?

- Key observation: each key pressed is transmitted separately.
- Depending on the position of the key on the keyboard, different inter key timings.
- Attack (*Song et al.*): observe the inter key timings (many times if you wish) – infer what keys have been pressed.
- Result: reduce the entropy of password – fewer guesses required.

Note that there is still variability across different people. Adds noise – but also opportunities (*Rubin et al.*)!

- Monitor a user session and record the timings of key presses.
- Use existing profiles to infer their identities according to the leaked timing.

Can extract both information and identity from a 'secure' session.

Do Secure Sockets (SSL/HTTPs) protect your privacy?

SSL is used to 'hide' sensitive web information (HTTP encrypted and authenticated) – but does it hide everything? (*Hintz et al, Simon et al.,. . .*)

- HTTP retrieves many resources per request (HTML page, style, images, . . .)
- SSL does not disturb timing much – doesn't hide length well.
- Attack: profile the website using SSL. For each possible request make a list of retrieved resources and their lengths.
- Observe the sequence of retrieved resource lengths of the victim – make a (good) guess about which page their correspond to.

We do better if we observe a sequence of requests (*Danezis*).

- Note that users are most likely to follow links on pages.
- Try to guess not only one request but a sequence – can use hidden Markov models to do this efficiently.

Even if SSL is used web click streams can be revealed.

Can I guess which pages you visited before? (Without observing you!)

Have you visited my competitor's website before visiting mine?

- Adversary is a hostile website that tries to determine browsing behaviour.
- Cannot directly observe the victim. The victim only makes one request to the hostile site.
- Key observation: modern browsers have caches of pages visited – good for efficiency.
- A resource in the cache will load much faster than if requested from the network.
- Attack: embed in my website a sequence of pictures from my competitor's site. Note how long it takes the browser to load these resources. Estimate if they were in the cache. Bingo!

Anonymizing proxies do not help! (Attack by *Felten et al.*)

Identification – are two network hosts the same machine?

How do I know if two different network addresses are the same machine? (*CAIDA*)

- Key observation: clock crystals have a variable drift – sensitive to heat conditions.
- If I can measure the time (ICMP, TCP time, web, . . .) I can estimate the drift.
- If over a period of time the drift matches it is the same machine.
- Applications: estimating number of consolidated servers, honey pot detection.

A second attack: crystals are sensitive to heat – can use that to find their location (time zone) by the day night cycles (*Murdoch et al.*). If the attacker has access to weather stations their might do better!

Identification – is one network host many machines?

A single NAT gateway or firewall can 'hide' behind it many network hosts. How many? (*Bellovin*)

- Need a way of differentiating different hosts from the traffic the gateway relays.
- Key observation: many TCP/IP network stacks implement the IPID as a simple counter (Windows). Every time a packet is generated it is increased by one.
- Attack: Observe all TCP/IP packets from the host, and plot their IPID numbers over time. Fit plausible straight lines – their number is the unique hosts.

Field of network mapping and network measurements. Attack tools like *nmap* available and *very sophisticated* (indirect port scanning).

Detecting stepping stones

Traffic analysis can be used for intrusion detection (defence).
Problem: I want my firewall to detect whether any host in my network is in fact compromised and used to relay attack streams.

- Firewall observes all incoming and outgoing TCP/IP connections.
- Their contents may well be encrypted (particularly if used by attacker).
- Passive detection: Use inter-packet delays and bounds on incoming streams latency to find out relays.
- Active detection: establish pseudo-random inter-packet delays (watermark stream) in incoming streams and try to detect them in outgoing.

Some assumptions are dodgy – but still interesting work.

Location information

Traffic data from cellular/GSM phones, WiFi base station registration and DF can be mined. Results from early studies:

- *Pascual et al.* studied WiFi access point data at HAL and KTH. Could infer talks/lectures attended by owner of machine. Could infer relationships by common patterns of movements.
- Intel Cambridge ran a bluetooth discovery experiment. Devices would record what other devices they see. Two members of staffs' devices were seeing each other at night.
- MIT Reality Mining: 100 Media Lab staff and students were given mobile phones and traffic data was recorded. Could infer friends (Saturday 8pm), could infer status (entropy of location), could predict movements.

Location data can be used to infer movements, relationships, status, . . . not just location!

Inferring high level information (I)

So far we have attacked specific security properties – lack of imagination! We can use traffic analysis to get more high level information about computer and human networks.

- Sociology and peer-to-peer networking theory is merging.
- Sociology: human networks have a low degree and are navigable (you can efficiently route to anyone with only local information!).
- For this you need a certain degree distribution (power law), or distribution of contacts.
- Resilience: very resistant to random failures.
- Weakness: fragile against targeted attacks.
- Can use traffic analysis for target selection! (high degree or betweenness)

Inferring high level information (II)

(Getting into policy. . .)
Work in making resilient covert networks / uncovering them.

- Use of social network analysis in criminal investigations. Lessons about who to arrest and who to monitor further.
- Resilient topologies (*Nagaraja and Anderson*). Make power law network when things are calm, random when under attack.
- The evolution of resilient peer-to-peer systems, under the constant threat of the copyright holders. The 'criminal mass' is pushing for the creation of more covert communications that are resilient to traffic analysis and infiltration (is this a good thing?)

/bf Mention in counter-insurgency doctrine manual!

Using traffic analysis for good, not evil

Traffic analysis inspired techniques can also be used to build *robustness* and *trust*.

- Google's PageRank is using contextual information (links amongst pages) to establish authoritative pages.
- Advogato – a social network, where free software developers are meeting. Introductions allow for very effective spam filtering.
- Relationship information (that can be inferred from traffic analysis) can be used to defeat the sybil attack in distributed peer-to-peer systems.
- Detect communities, hubs and authorities about a field or subject.

Invaluable in information retrieval, ambient intelligence, ad-hoc networks but also price discrimination and profiling.

Traffic analysis resistance

Over 20 years of research but only recently very active.

- Anonymous communications – hide link between senders and receivers.
- Location privacy – reduce the resolution of traffic data / linkability.
- Censorship resistance (jamming resistance): peer-to-peer, firewall piercing (China or VoIP), steganography, cover channels, ...
- Spam filtering, abuse resistance, popup blocking, (deception resistance).
- More generally: Privacy Enhancing Technologies (credentials, minimum disclosure, privacy policies, database inference, ...)

Anonymous communications – a primer

Theoretical / research time line

- 1981 – David Chaum introduces the mix.
- 1985 – DC netowrk / unconditional anonymity
- 1991 – ISDN Mixes (*Pfitzmann*).
- 1996 – Onion routing.
- 2000 – Freenet (still not quite right) & p2p

Fielded systems

- Remailers for anonymous email: Penet (88), cypherpunk (92), Mixmaster (94), Mixminion (03).
- Web anonymity: Jap (01), Tor (03), (Anonymizer 95?)

Recently: a lot of research on analysing and understanding anonymous communications. Qualitative shift in the attacks and defences.

The traffic analysis of anonymous communications

Remailers

- Slow delivery, inflexible size of packets.
- Very secure against global active adversary.
- Long term intersection attacks still possible to infer long term relationships.
- Weak against infiltration (still)

Onion routers

- Low latecy for interactive streams.
- Weak against even passive adversaries.
- Correlation attacks are possible to trace streams.
- Assume weaker threat model (local attacker)

Sophistication of attacks is astonishing: Tor was attacked with no access at all to any traffic data but your own! (*Murdoch and Danezis*) Heat! (*Murdoch*).

Key policy issues: Traffic data retention

What is traffic data retention

- E.U. and G8 are implementing traffic data retention.
- Certain categories of traffic data kept for years.
- To facilitate future investigations.
- ISPs / Mobile providers to bear the cost of storage, access and security.

What are the issues

- Introducing a *systemic risk* of exposure to traffic analysis.
- Covert communication networks can be established despite even the most stringent retention regimes.
- Extent to which these data can be used for attack is understated.

Key policy issues: Peer-to-peer information warfare

Co-evolution of peer-to-peer file sharing with information and legal attacks.

- Music swapping networks are using more and more covert techniques – dark nets, traffic analysis resistance, anonymous communication architectures, minimum disclosure, spamming prevention, ...
- Copyright holders are using increasingly hostile measures: traffic analysis and target selection for legal action, disruption of indexing and distribution, spamming of networks, deception, ...
- At the same time the peer-to-peer paradigm is catching on: VoIP (skype), BitTorrent are worth billions.
- Is this constant tussle hampering network, technology and business model innovation? Are we hurting the IT industry in order to keep the music industry alive? Is it beneficial to have an industry based on attacking these p2p networks?

Key policy issues: Privacy and data protection

Traffic analysis is a key threat to the right to privacy.

- Traffic data information can be used to infer highly private attributes.
- Mining increases its value in a way that is inconceivable for non-specialists.
- Definitions of private information (data protection) fall short of protecting against information refining.
- Not enough support is provided to deploying PET technologies.
- Mining on the other hand is well funded, and systematic.
- Privacy invasion is like pollution: the more you know about my neighbour the more you know about me! Yet no one has direct strong incentives to act.

Conclusions

- Traffic analysis has been neglected for too long by the mainstream security community, despite being of vital importance when it comes to operational security.
- Lessons and paradigms from the military world teach us about techniques that can be used to attack civilian networks and security policies.
- The level of sophistication of the attacks, and defences is advanced – established body of knowledge in the open community should not be ignored.
- Secure communications are still expensive – and often fragile.
- Policy decisions that minimise exposure to attack even more important – the opposite is often observed.
- This, along with other network attackers, further fuels the deployment of covert communication networks.

More, more, more...

- We are **looking for students/researchers** at K.U.Leuven.
- The Privacy Enhancing Technologies Workshop (PET2007), Ottawa, Canada, May/June 2007.
  `http://petworkshop.org`
- Anonymity bibliography: `http://freehaven.net/anonbib/`
- Contact me: `gdanezis@esat.kuleuven.be`
  (or Roger Dingledine, or Steven Murdoch)