

WiFidog; running a public wireless network in Montreal

- Benoit Grégoire, Technologies Coeus inc.

Why am I here?

- Data from public networks is hard to come by, most companies consider that dangerous strategic information...
- A dataset from a real public network in Montreal
 - Covers all connections on the Île sans fil network from 2004-08-27 to 2007-08-28
 - Anonymised SQL format

Who is Île sans fil?

- A Non profit organization entirely run by volunteers
- First hotspot went live 2003-07-15
- Current mandate
 - Providing free Internet access to get citizens in public places
 - Disseminate location-relevant art and community content

Description of the network

- Network of WiFi hotspots in public spaces
- Independent hi speed connections
- Central content delivery and authentication
- Users have to open an account with a persistent email address
- Connection speed is not capped
- The largest Hotspot network in Montreal
 - 47 000 validated users
 - 130 Hotspots

Curent business model

- Internet access is free for users
- Connectivity is provided by the venue owner
- Venues pay a 50\$ yearly membership to join ISF, and (typically) 100\$ for the hardware.

Facing the challenges: Equipment cost

- Standardized on the Linksys WRT54G(L)
(~80\$)

Facing the challenges: Equipement failure

- Overall, has proven extremely reliable, except for one revision (2.2)
- Staff just won't leave equipement alone...

Facing the challenges: "Business" hi-speed is unreliable

- Yes, DSL modems are cheap, but they aren't the worst problem!
- DSLAMs go down
- Entire backbones go down (more frequently than the telcos would like to admit)
- Phone lines conditions are horrible, and other technicians tinker with them

Facing the challenges. supporting a network you don't control

- Placement of equipment is critical.
- We tend to take the blame anyway...
- Wifidog is completely NAT-in
- Modems aren't the worst problem
- DSLAMs go down
- Entire backbones go down
- Phone lines are horrible, and other technicians tinker with them

■ We tend to take the blame...



Facing the challenges: Making deployment easier

- A single, self-configuring firmware image.
 - <http://www.ilesansfil.org/FirmwareISF>
- A smarter auth server

What is WiFiDog

- A content oriented captive portal system.
- A network management platform
- The WiFiDog was initiated by Île sans fil, to replace NoCat
- Developpement direction and speed are determined by the resources spent by it's community

Design goals

- Architectural flexibility
- Supporting all of <http://dev.wifidog.org/wiki/WirelessCommunityModels>
- Not imposing our policy choices to other groups
- Architecture
 - Ultra-thin client with smart central server

Location specific content

- The problem: There are several kind of location specific content.
 - "about the place" ex: Historical information, price list, etc.
 - "at the place" ex: Events
 - "related to the type of place or it's community community" ex: The owner's personal blog,
- Maybe it should be an editorial decision remembered by wifidog
- Little geolocated content is available

Location portal UI are difficult to agree on

- Location portals are a brand new media, and are poorly understood.
- We won't get anywhere if people keep seeing them as location-specific website, or as a network-wide broadcast platform

WiFiDog Content delivery

- The portal is unique
- Content elements can be eligible for delivery at a node, and for a specific user according to specific criteria
- Many different content types (Text, HTML, Images, but also Content groups, Smarty templates, Stylesheets, etc.)
- Language-aware content delivery

WiFiDog Content delivery: Temporal criterias

- Content can be eligible for delivery
 - starting at a certain date
 - untill a certain date
- Content rotation can be:
 - Systematic
 - Once per day
 - Once per connection
- Grouping
- Can be influenced by past events

WiFiDog Content delivery: Locative criterias

- Content can be eligible to be delivered at a single node or group of nodes
- Content rotation can be triggered by physically changing node
- Content can be eligible for delivery only once at a single node

WiFiDog Content delivery: RSS feeds

- A very sophisticated aggregator, designed to pick the "freshest" content from a group of news feed, like a human would
- Also support podcasts
- Does not support GeoRSS yet

WiFiDog Content delivery: The trouble with RSS

- The quality of RSS feeds is degrading
 - Apple iPhoto "standard"
 - Flickr doesn't even have photocasts available
 - Re-aggregation of Youtube (youtube itself is fine)
 - People manually typing HTML
- How did we get there?
 - People tend to only care how a their specific feed displays at a specific site
 - The general problem with metadata, is that it only has value if many people use it, in mostly the same

Data collection

- The actual data collected is extremely limited, for privacy reasons:
 - account id
 - MAC and IP addresses
 - login and logout time
 - hotspot
 - amount of data transferred.
 - content display and clickthrough, but only for content we display on the auth server.

Data analysis

- Having unique users makes this data extremely rich
- WiFiDog already has a pretty sophisticated reporting framework, with several built-in reports
- Data analysis is on-demand

Report engine

Report configuration

Network: Île Sans Fil

Restrict the time range for which statistics will be computed :

From:

Select from and to...



To:

Select from and to...



Restrict stats to the following nodes :

Arts Café
Atomic Café / Le 7ième
Atwater Library
Auberge Alternative du Vieux-Montréal
Bar Baloos
Bar Chez Baptiste



Distinguish users by :

Username



Restrict stats to the selected users :

Username or MAC address, depending on selection above

Selected reports :

- 10 highest bandwidth consumers
- 10 most frequent users
- 10 most mobile users
- Anonymised SQL data export (for academic research)
- Breakdown of how many users actually use the network
- Connection Log
- Content display and clickthrough report
- Graph on network use per hour, weekday and month
- Individual user report
- Most popular nodes, by visit
- Network status information
- Node status information
- Registration Log (New user's first connection)
- User registration report

Generate statistics



So how many users use the network

Breakdown of how many users actually use the network

User activity

Activity report for the 47057 validated users

Period	# of users who used the network
Last day	649
Last week	2084
Last month	5276
Last 3 month	10052
Last 6 months	16780
Last year	25247
Ever	38960

Activity report for the 36761 non-validated users

Period	# of users who used the network
Last day	94
Last week	429
Last month	1492
Last 3 month	3795
Last 6 months	9111
Last year	19718
Ever	32893

warning: This report does not count connections at Splash-Only nodes

So how many users use the network, by MAC addresses

Breakdown of how many users actually use the network

User activity

Activity report for the 47058 validated users

Period	# of users who used the network
Last day	670
Last week	2183
Last month	5558
Last 3 month	10442
Last 6 months	17297
Last year	25628
Ever	38929

Activity report for the 36764 non-validated users

Period	# of users who used the network
Last day	64
Last week	292
Last month	908
Last 3 month	2284
Last 6 months	4750
Last year	8495
Ever	14167

warning: This report does not count connections at Splash-Only nodes

A few gotchas

- Connection count is meaningless as a metric
- Online time is interesting, but of limited reliability

Analysing content delivery (not part of CRAWDAD dataset)

Content display and clickthrough report

Select from reusable content library:

Survey (HTMLeditor)

Add

Content report for: Survey<:

	Clickthrough	Prints	Clickthrough/Prints
Count	625	55365	1.13%
First	12/13/2006	12/13/2006	
Last	04/30/2007	04/30/2007	
Rate	4.53 per day	400.86 per day	
Unique users	566	9367	6.04%
Unique locations	109	161	67.70%

Important note: Currently, Report configuration options are ignored for this report.

What we'd like to know: Clustering

- Finding groups of users using different hotspots.
- Watch out to compensate for hotspot popularity
- If we run you analysis on the real data and we don't find the active ISF volunteers having meetings, we'll probably conclude that your methodology is flawed

The future: Dynamic abuse control

- Abuse control was one of the originally stated goals of WD, but has been put off
 - We have a very good scaling model
- Why should you do abuse control in the first place
- Past solutions
 - Participating in the arms race is silly!
 - Solutions focussed either on ISF, or on running Bittorrent and WoW at the same time.