

“E-mail Forensics”

Les Hatton

CISM, Kingston University
L.Hatton@kingston.ac.uk

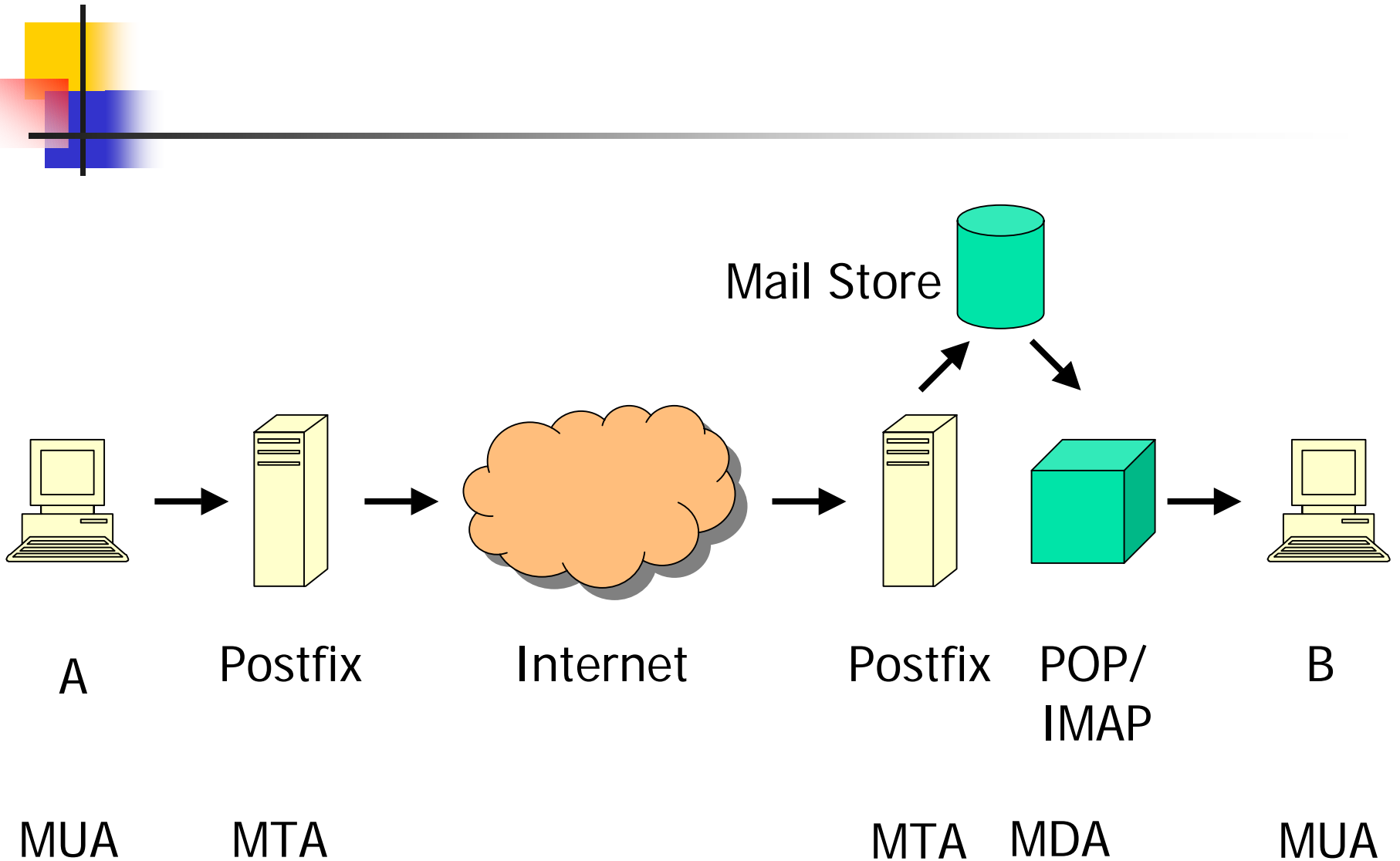
Version 1.1: 26/Mar/2010

Overview



- Overview
 - The basics
 - The threat landscape
 - Commercial and non-commercial filtering
- Defence in Depth
- Wrap-up

The basics: A → B



The basics: Headers and Content

Envelope

```
Connect mail.receive.com (an MTA responds)
HELO mail.send.com (an MTA responds)
MAIL FROM: alice@send.com (an MTA responds)
RCPT TO: bob@receive.com (an MTA responds)
DATA (an MTA responds)
```

Content

```
Date: ...
From: alice@send.com
To: bob@receive.com
Reply-To: ...
Message-ID: ...
Subject: ...

Blah blah blah
```

```
Disconnect
```

MTA



MUA



The basics: Headers and Content

Send
MTA



Connect mail.receive.com (an MTA responds)
HELO mail.send.com (an MTA responds)
MAIL FROM: alice@send.com (an MTA responds)
RCPT TO: bob@receive.com (an MTA responds)
DATA (an MTA responds)
....



Receive
MTA



Accept
(loses HELO)

Reject or Discard via HELO, ...



Backscatter ! Bounce via From:



*** Reject as early as possible ***



The basics: what can be forged ?

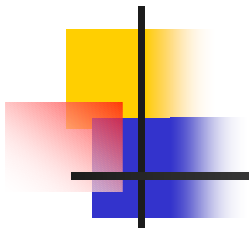
- Headers that can be forged
 - Subject, Date, Message-ID
 - Recipients: From, To, CC, BCC
 - Content body
 - Any arbitrary headers, X-Mailer ...
 - All but the last Received header
- Headers that can *not* be forged
 - Last (top most) Received header
 - Originating mail server, specifically
 - IP address
 - Subsequent timestamps

The basics: anybody can read it



- Nearly all e-mail is sent in clear as if you had written it on a postcard and asked a complete stranger to post it for you.
- It can be arbitrarily spoofed

The basics - things you need to know



- Perl and regular expressions
 - Header checks, body checks and content filtering
 - SpamAssassin
 - Basic data analysis
 - Postfix
 - SMTP
- and a little about ...
- DNS
 - TCP/IP

The basics – Perl



- CPAN available plugins for parsing e-mail messages <http://www.cpan.org/>
 - MIME and attachments
 - HTML
 - Looking for toxic links
 - SQLite
 - Geographic IP address validation
- Know your regular expressions, for example,
 - `$url =~ m|(\w+)://([^\:]+)(:\d+)?/(.*)|;`
 - `$protocol = $1, $domain = $2, $port = $3, $uri = $4.`

The basics – SpamAssassin



- A number of important features
 - Grades spam by summing various contributing factors (hundreds of rules). Anything over 5 is almost certainly junk.
 - You can add your own rules,
<http://www.spamassassin.org/>

The threat landscape – e-mail borne toxins



- Spam
 - Density
 - Works of Art
 - Harvesting
- Scams
- Intrusions

The threat landscape - spam

- For a single mail-server handling mail for 8 domains in 2-9 November

Total received	836,106	100.00%
Discarded	217,275	25.99%
Rejected	618,000	74.00%
Rejected by content filtering	13	0.002%
Delivered to users	818	0.098%
(Missed spam / lost mail)	(0/0)	0.00/0.00%

The threat landscape – HTML works of art

```
<body>=09
```

```
<p>=09What's up?<a name=3D"#tprw"></a></p><a name=3D"#qpqr"></a><span name=3D"#twqp">=09</span><br><a name=3D"#rtwp"> </a><table border=3D="6" cellspacing=3D="7" cellpadding=3D="1" width=3D"199">
```

```
<tr><td bordercolor=3D"#4B41CE" nowrap=3D"nowrap" valign=3D"baseline" bgcolor=
=3D"#D9F0BC"><strong>V</strong><font color=3D"#D9F0BC">b</font> </td>
```

```
<td nowrap=3D"nowrap" valign=3D"middle" bordercolor=3D"#69FA49" bgcolor==3D"#BCB6F0"
align=3D"center"> I </td><td bordercolor=3D"#67DA87" valign=3D"baseline" bgcolor=3D"#F0BCC3" align=
=3D"left" nowrap=3D"nowrap"> <b>A</b></td>
```

```
<td align=3D"center" bgcolor=3D"#F0C3BC" bordercolor=3D"
<td nowrap=3D"nowrap" bgcolor=3D"#D4F0BC" bordercolor=3D
color=3D"#D4F0BC">y</font> </td>
```

```
<td align=3D"left" bordercolor=3D"#6852DC" valign=3D"top"
<font color=3D"#F0B9BC">3</font>A=09</td></tr></table>
```

```
=09<br><strong></strong><table><tr><td>WWW</td><span>
</td><br><td>.</td><span name=3D"#rrtp"></span><td>COM</
</b><br><strong>=09</strong><p><span></span></p><span na
name=3D"#qqtp"></a></p>
```

```
</body>
```

From: "Riso Nuzzi" <cohered@nda.co.nz>
To: gundalf@oakcomp.co.uk
Date: 2008-08-07 00:10

Spam Status: Spamassassin

Heya,



WWW.NEVOB.COM

The threat landscape – harvesting e-mails



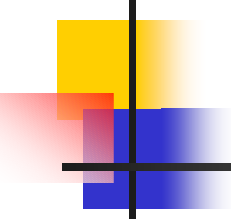
- “New virus coming – warn 25 of your friends
...”
- “New speed camera – pass on to your friends”
- Assorted nonsense of a similar kind.

The threat landscape – e-mail borne toxins



- Spam
- Scams
 - Nature
 - Density
- Intrusions

The threat landscape – Nature of scams

- 
-
- Of various kinds
 - Lottery (often the Netherlands)
 - “My left leg has been bitten off by a mad cheetah and I have \$4 million in da trouser leg” (Nigerian 419)
 - Phishing for account details
 - Pharming, DNS hijacking, ...

The threat landscape – Density of scams



- Typical density
 - Approximately 2 per day per domain name but growing

The threat landscape – e-mail borne toxins

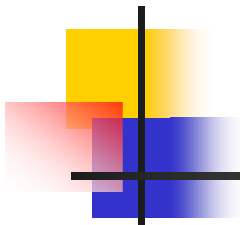


- Spam
- Scams
- Intrusions
 - Click-throughs
 - Attached files

The threat landscape – Intrusions

- Click-throughs:
 - Encourage clicking on a link to install a trojan or bot
 - Encourage you to hand over personal details
- Attached files
 - Encourage clicking on a zip file with the subject, “Your invoice”, “Your receipt”, “Your ...”

The threat landscape – recent example



Came from compromised mailbox in KU. User does not exist.

From: address does not exist

Link –zyef.9hz.com does not exist


Your mailbox is almost full.

Brannan, Carine J

Sent: Tue 23/03/2010 12:48

To: info@webmailhelpdesk.org

Your mailbox is almost full.

20GB  23GB

Your Webmail Quota Has Exceeded The Set Quota/Limit Which Is 20GB.
You Are Currently Running On 23GB Due To Hidden Files And Folder On Your Mailbox.
Please Click the Link Below To Validate Your Mailbox And Increase Your Quota.

[Click Here](#)

Failure To Click This Link And Validate Your Quota May Result In Loss Of Important Information In Your Mailbox/Or Cause Limited Access To It.

Thanks
HELP DESK

Commercial and non-commercial filtering



■ Commercial

- Buy tools
- Buy service, (eg MessageLabs)
 - Your mileage may vary here. ML have missed an average of 15 a day in my mailbox since 6th September, mostly backscatter, with some days hopeless.
- Use Google, yahoo or somebody
 - Be warned. All e-mail is read by tools for filtering. It would be easy to store keywords with other information already held on you.

■ Non-commercial

- Tools like SpamAssassin

Overview



- Overview
- Defence in Depth
 - Layered protection and Postfix
- Wrap-up

Layered protection and postfix - essentials

- Reject as early as possible in the transaction
- No open relays

Connect mail.receive.com (an MTA responds)

HELO mail.send.com

(an MTA responds ...)

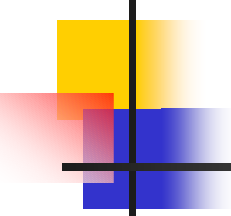
MAIL FROM: alice@totallybogus.com

250 Ok

RCPT TO: bob@notonyourserver.com

554 bob@notonyourserver.com Recipient address
rejected: Relay access denied

Layered protection and Postfix



smtpd_client_restrictions

smtpd_helo_restrictions

smtpd_sender_restrictions

smtpd_recipient_restrictions

smtpd_data_restrictions

header_checks

body_checks

external content filtering

Server 220 smtp.receive.com ESMTP Postfix

Client: HELO mail.send.com

Server 250 smtp.receive.com

Client: MAIL FROM: alice@send.com

Server 250 OK

Client: RCPT TO: bob@receive.com

Server 250 OK

Client: DATA

Server 354 End data with <CR><LF>. ...

Client: To: bob@receive.com

From: alice@send.com

Subject: example

Message body, blah blah, blah

Other content – Bayesian etc.

Layered protection and Postfix – typical installation with latest volumes



smtpd_client_restrictions

(618,000) smtpd_helo_restrictions

(neg.) smtpd_sender_restrictions

(216,000) smtpd_recipient_restrictions

(neg.) smtpd_data_restrictions

(~ 200) header_checks

(~ 200) body_checks

(~ 100) external content filtering

None

reject self_helo, reject_non_fqdn_hostname,
reject_invalid_hostname

reject_non_fqdn_sender,
reject_unknown_sender_domain, blacklist

Reject_non_fqdn_recipient,
reject_unknown_sender/recipient_domain,
reject_unauth_destination, reject unknown users,
reject_rbl_client (spamcop etc.), **Greylist**, back
scatter

Reject_unauth_pipelining

Different from/return-to, foreign character sets,
various Windows skullduggery

Other Windows skullduggery, tags, attachments,
embedded .exe, ...

Viral filtering, sigs, creative html, SA, domain
mismatch, embedded SURBL, cross-domain match



Layered protection and Postfix

- About 75% of all e-mails can be rejected by being unable to say HELO properly.
- If you filter properly, hardly any viral attacks get through, (on my own mail servers, its around 1 per month out of a total of around 5 million e-mails)
- The best RBL is unquestionably Spamhaus, but SpamCop is also good.

Last line of defence – content filtering



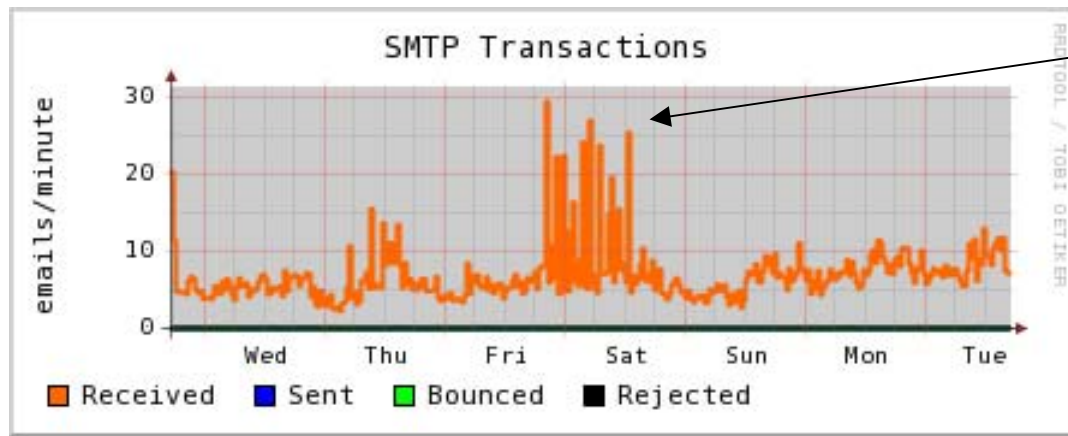
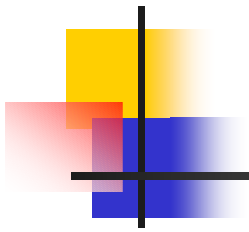
- Some things you can implement if your Perl is up to it
 - MIME parsing and analysis of attachments.
 - HTML parsing and identification of embedded `<script>`, ``
 - More complex dictionary checks than you can do easily with SpamAssassin.
 - Geographic lookup of envelope or embedded IP addresses

Miscellaneous

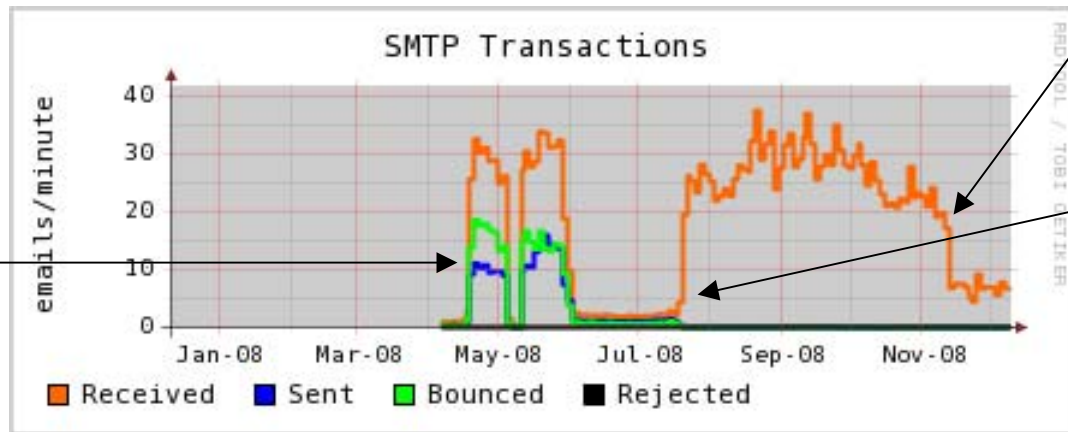


- Dealing with backscatter
- Unsubscribing
- Honeypots
- Whitelisting (as a last resort)

Learning on the job



Typical week



Year to date

LH failing to understand the FormMail relay injection loophole

Weekend rubbish

Attivo site closed in USA

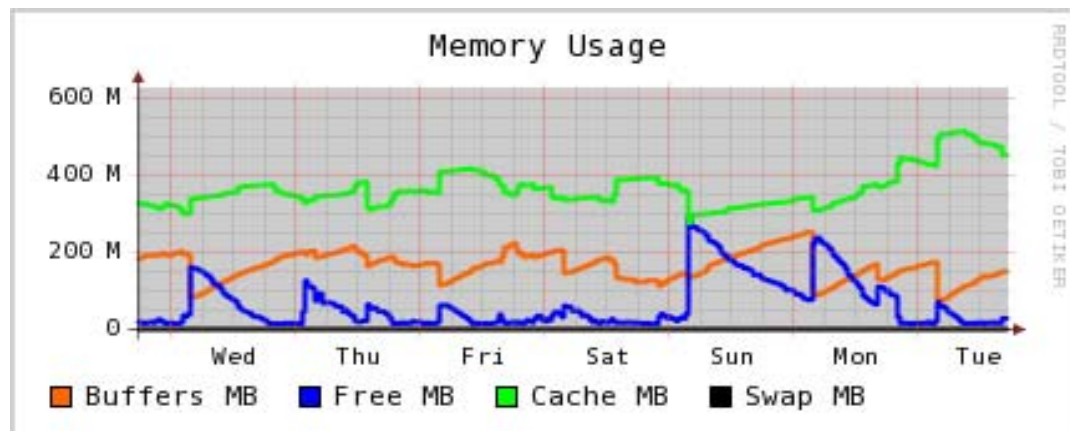
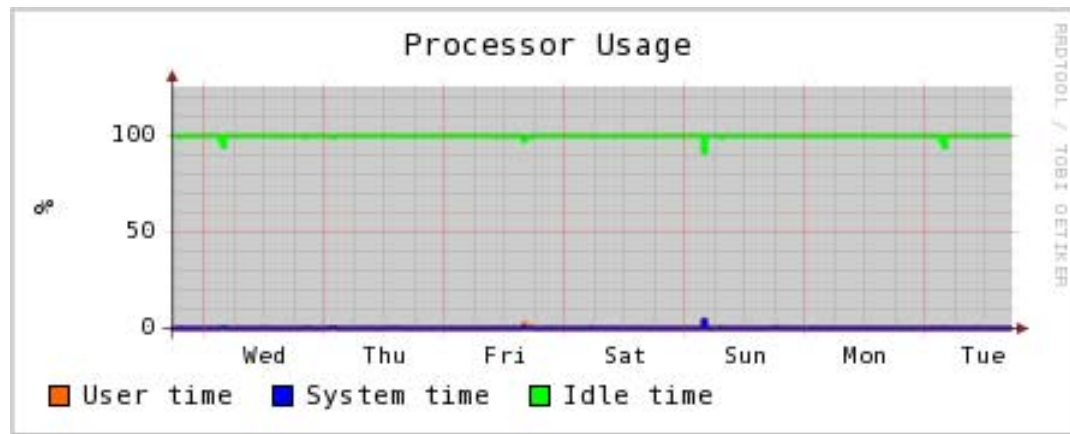
Silent discard, greylisting, RBL

Overview



- Overview
- Defence in Depth
- Wrap-up
 - Load on server
 - Threats in the pipeline
 - Things you need to know
 - Things still up our sleeve

Load on server



Threats in the pipeline



- Scamming activity and identity theft is up and getting more sophisticated – be careful !
 - 2009-10 Significant escalation
 - In 2007 3.6 million adults *claimed* to have lost US \$3.2 billion.
 - In 2005, around GBP 24 million was lost in phishing.
- Lots more opportunity with hand-helds.
- Bot nets still growing sadly due to user ignorance and Windows vulnerability (~ 25% of all PCs)
- Pharming – DNS poisoning; watch your router.

Threats in the pipeline: botnets



- Networks of PCs (20,000 – 500,000) controlled externally, often by IRC (Internet Relay Chat) servers.
- How they work
 - Botnet operator (herder) sends out viruses or worms infecting ordinary users' PCs.
 - The bot connects with the IRC server
 - Spammers purchase access to a botnet



Things still up our sleeve

- SPF
- Challenge / Response
- Lots of things we can do in content filtering

Conclusions



- It is possible to operate as near to zero-spam as makes no difference
- Most spamming is still unsophisticated
- Scamming, (phishing, pharming) is getting much better and is a real danger rather than simply being a nuisance
- This will become more of a challenge

References



Loads of stuff on Wikipedia:-

<http://www.wikipedia.org/>

My writing site:-

<http://www.leshatton.org/>