

# Talkback: Reclaiming the Blogosphere

Elie Bursztein, Baptiste Gourdin, John Mitchell  
Stanford University & LSV-ENS Cachan

# What is a blog ?

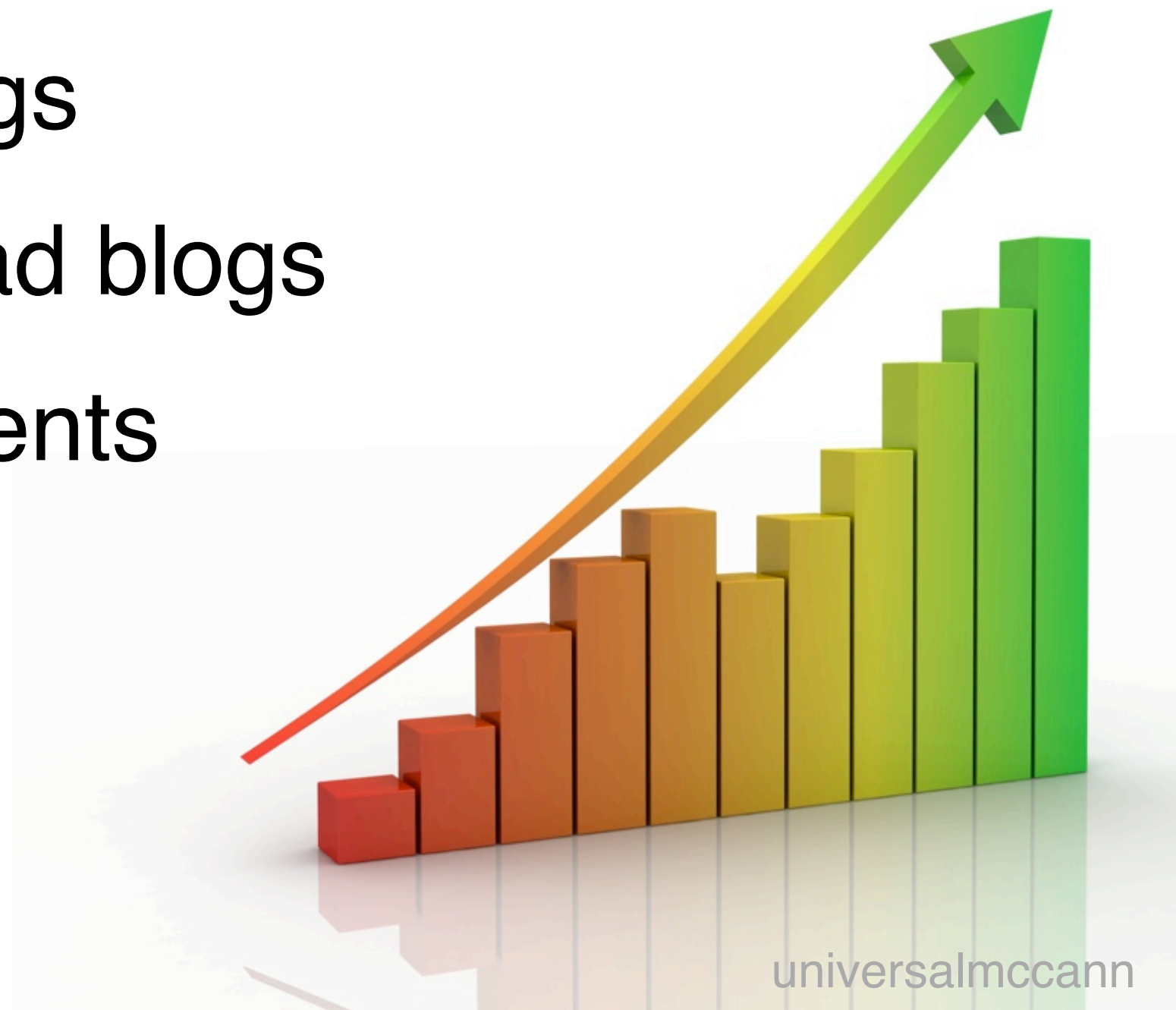
- A Blog ("Web log") is a site, usually maintained by an individual with
- Regular entries
- Commentary
- LinkBack
- Entries displayed in reverse-chronological order.



<http://elie.im/blog>

# Key Statistics

- 184 Millions blogs
- 73% of users read blogs
- 50% post comments





# Anatomy of a blog post

The image shows a screenshot of a WordPress blog post editor interface. The interface is divided into several sections, with numbered annotations (1-7) highlighting specific elements:

- 1. Title:** The title field contains "New MailChimp Logo By Jon Hicks".
- 2. Permalink:** The permalink field shows the URL: <http://www.mailchimp.com/blog/new-mailchimp-...o-by-jon-hicks/>.
- 3. Body / Content:** The main content area contains text: "This post has absolutely *nothing* to do with email marketing. But if you're interested in graphic design and branding, you might enjoy..." and "I've been wanting to redesign the MailChimp logo. I first designed it in 2001. I kinda hacked the logo together really fast in Fireworks (which is just not meant for high-res print). I think it went something like this. Mark, my co-founder, said: 'Yo Ben, MailChimp's logo is terrible. It could use a logo.' My 'Oh yes, here, this sketch...'". There is also a small image of a cartoon character holding a sign that says "MailChimp".
- 4. Publish Date:** The publish status is set to "Published". The publish date is "September 11, 2008 at 4:49 am".
- 5. Category:** The category selection area shows "All Categories" and "Most Used". The "IMHO" category is selected.
- 6. Summary / Excerpt:** The excerpt field contains the text: "The story behind MailChimp's new branding, including sketches from Jon Hicks".
- 7. Author:** The post author is set to "Ben".

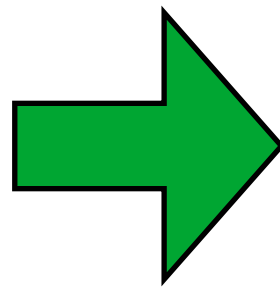
Other visible elements include the "Visual" and "HTML" tabs, the "Add media" button, the "Format" dropdown, the "Save" and "Delete post" buttons, and the "Related" and "Shortcuts" sections on the right.



# Why blogs are special ?



User



**Elie Bursztein** Web, mobile and offensive security research

Home About me Publications Talks Software Blog

### Social news

29 minutes ago RT @jeremiahg: "Expression Language Injection" ... "new, specific attacks against Spring MVC JSP tags" <http://t.co/unx26Vn> #security #java

about 3 hours ago Checking out <http://t.co/kppE3NQ> My summary of last week security and privacy top headlines is out #security #privacy

3 days ago Checking out <http://t.co/61MLajV> Firesheep addon updated to exploit Google info leak: 40% of your click history can be stolen. #privacy #security #www

3 days ago Reading of the day: <http://t.co/SzuUGsV> How Researchers used Typosquatting to stole 20 GB of E-Mail From Fortune 500 #security #privacy

3 days ago Reply to @mvyonline Sorry :) My ipad new reader tends to remove the HTTP prefix sometime for no reason ... #EPICfail

### Blog Posts

about 4 hours ago Security and privacy top headlines September 4th ? 10th

6 days ago Security and privacy top headlines

17 days ago ...

### Selected recent publications

**Towards Secure Embedded Web Interfaces**  
by Baptiste Gourdin Chinmay Soman Hristo Bojinov Elie Bursztein  
Usenix Security 2011, San Francisco, USA  
Tags: web, embedded devices, offensive technologies

**OpenConflict: Preventing Real Time Map Hacks in Online Games**  
by Elie Bursztein Jocelyn Lagarenne Mike Hamburg Dan Boneh  
S&P 2011, Oakland, USA  
Tags: game, crypto, offensive technology

**The Failure of Noise-Based Non-Continuous Audio Captchas**  
by Elie Bursztein Romain Bauxis Hristo Paskov Daniele Perito Celine Fabry John C. Mitchell  
S P 2011, Oakland, USA  
Tags: captcha, web, offensive technology

### Selected recent talks

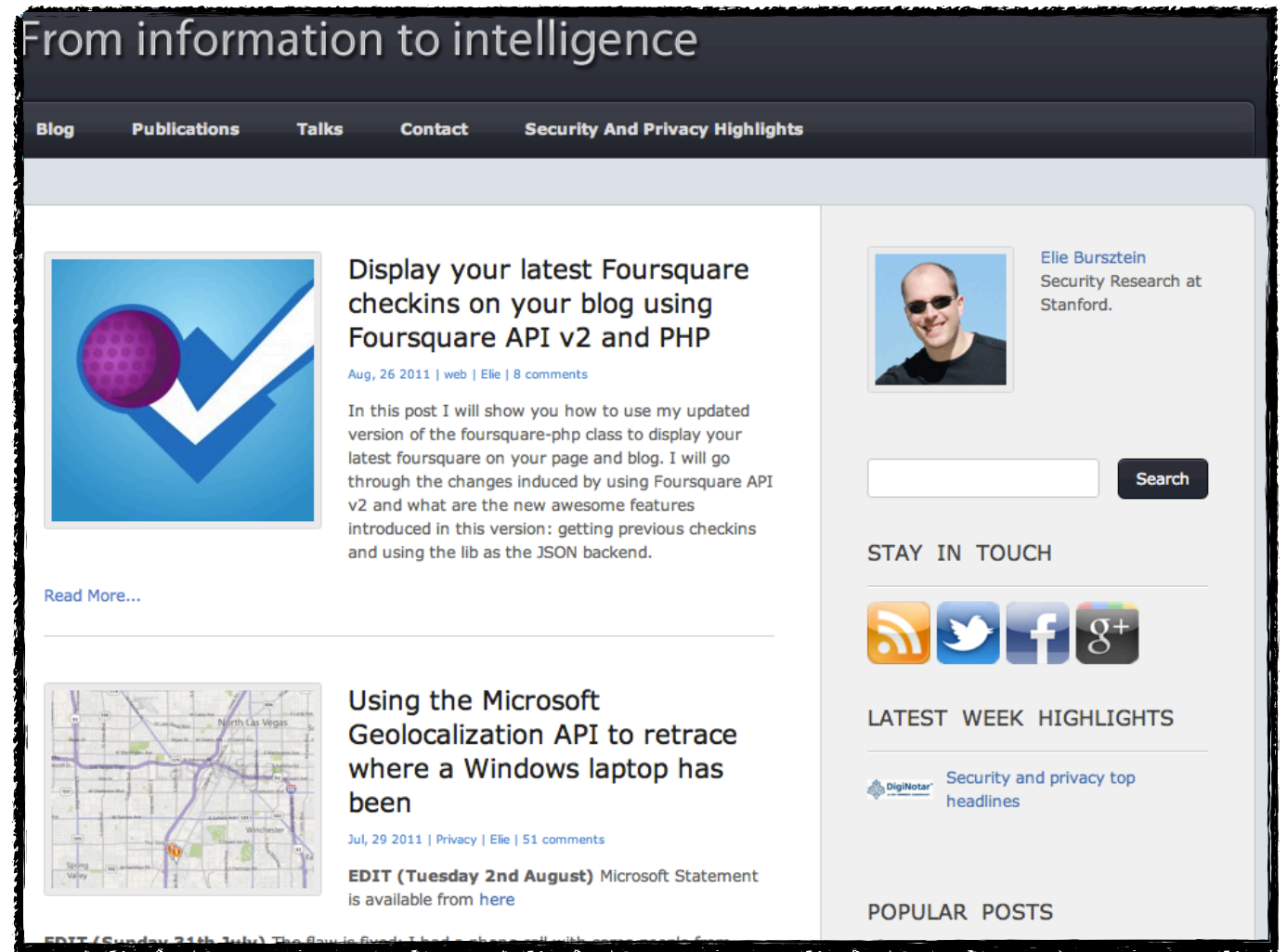
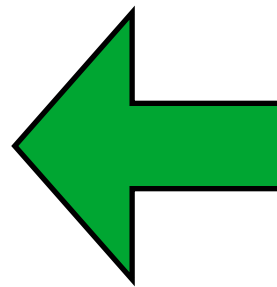
**Beyond files recovery, OWADE cloud-based forensic**  
BlackHat USA 2011, Las Vegas, USA  
Tags: forensic, windows

**Kartograph**  
Defcon 18, Las Vegas, USA  
Tags: game, offensive technologies

# Why blogs are special ?



User



# What is a TrackBack ?

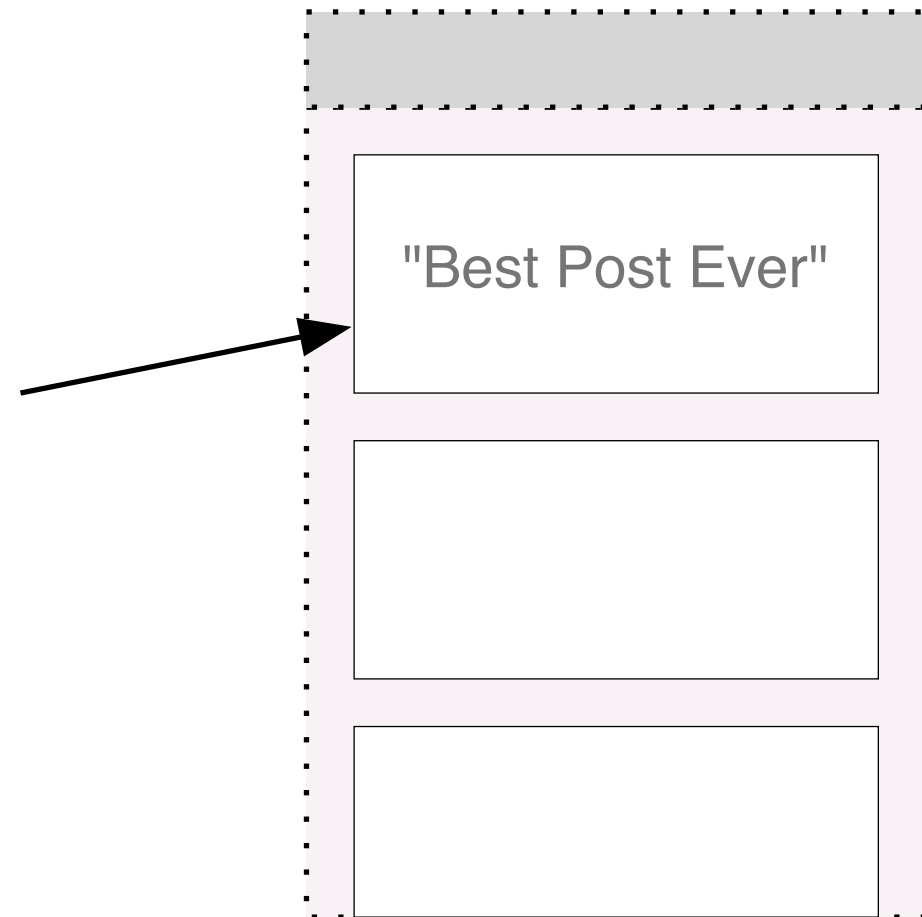




# Trackback Illustrated

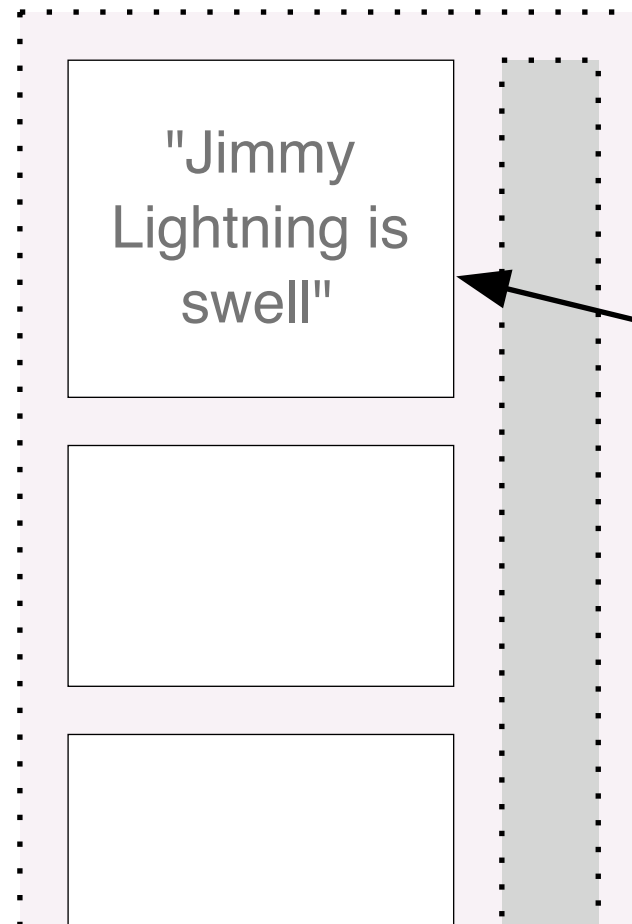
1. ■

"Wow! Jimmy Lightning has written the best post ever! It's so funny! And it's true! That's why it's so good. I need to tell the world!"



# Trackback Illustrated

2.



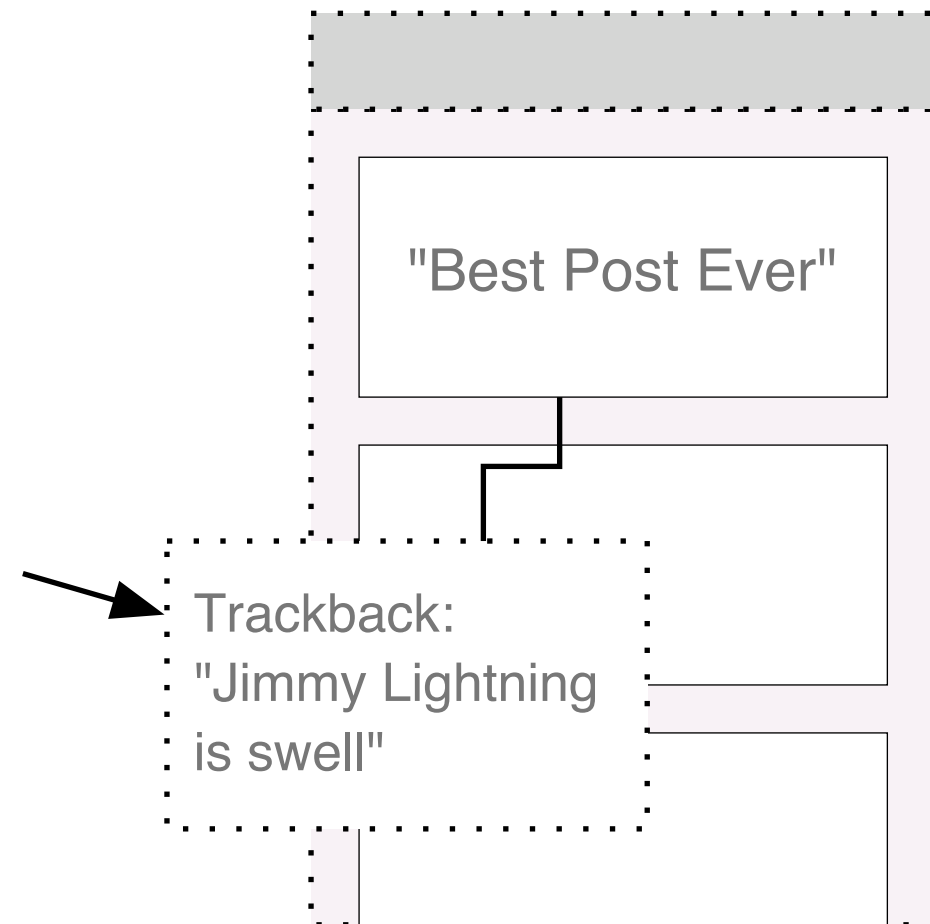
"Check it out world! I've written all about Jimmy Lightning's post on my weblog. My weblog's called '*The Unbloggable Blogness of Blogging*'. It's a good name huh? Wow, I sure hope Jimmy sees what I said about him..."

# Trackback Illustrated

3.

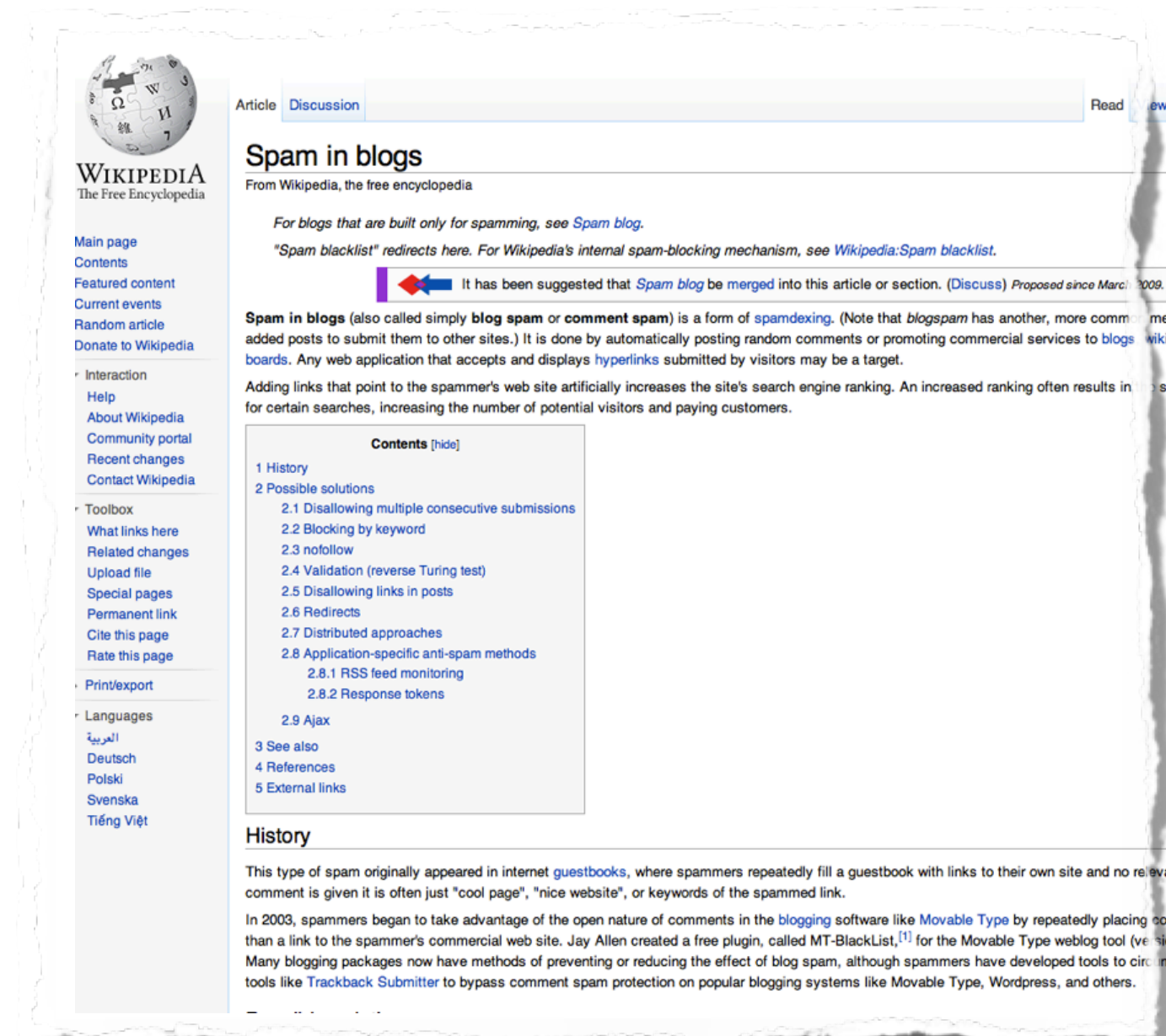
"Hey! I wonder what happens when I click on that Trackback link at the bottom of Jimmy's post?"

"Far out! In that funny pop-up is a link back to my site! Too cool for school!"





# Why the LinkBack problem is different ?



The screenshot shows the Wikipedia article "Spam in blogs". On the left is the standard Wikipedia sidebar with navigation links like "Main page", "Contents", and "Featured content". The article title "Spam in blogs" is at the top, followed by the subtext "From Wikipedia, the free encyclopedia". Below this is a notice about a suggested merge from "Spam blog". The main text defines "Spam in blogs" as a form of "spamdexing" and explains how it works by posting random comments or promoting commercial services. A "Contents" box lists sections: History, Possible solutions (with sub-sections like Disallowing multiple consecutive submissions, Blocking by keyword, etc.), See also, References, and External links. The "History" section begins by stating that this type of spam originally appeared in internet "guestbooks".

WIKIPEDIA  
The Free Encyclopedia

Main page  
Contents  
Featured content  
Current events  
Random article  
Donate to Wikipedia

Interaction  
Help  
About Wikipedia  
Community portal  
Recent changes  
Contact Wikipedia

Toolbox  
What links here  
Related changes  
Upload file  
Special pages  
Permanent link  
Cite this page  
Rate this page

Print/export  
Languages  
العربية  
Deutsch  
Polski  
Svenska  
Tiếng Việt


Article Discussion Read New

## Spam in blogs

From Wikipedia, the free encyclopedia

*For blogs that are built only for spamming, see [Spam blog](#).*

*"Spam blacklist" redirects here. For Wikipedia's internal spam-blocking mechanism, see [Wikipedia:Spam blacklist](#).*

 It has been suggested that *Spam blog* be merged into this article or section. ([Discuss](#)) *Proposed since March 2009.*

**Spam in blogs** (also called simply **blog spam** or **comment spam**) is a form of [spamdexing](#). (Note that *blogspam* has another, more common meaning: [added posts to submit them to other sites.](#)) It is done by automatically posting random comments or promoting commercial services to [blogs](#) or [wiki boards](#). Any web application that accepts and displays [hyperlinks](#) submitted by visitors may be a target.

Adding links that point to the spammer's web site artificially increases the site's search engine ranking. An increased ranking often results in more visitors to the site, increasing the number of potential visitors and paying customers.

**Contents** [hide]

- 1 History
- 2 Possible solutions
  - 2.1 Disallowing multiple consecutive submissions
  - 2.2 Blocking by keyword
  - 2.3 nofollow
  - 2.4 Validation (reverse Turing test)
  - 2.5 Disallowing links in posts
  - 2.6 Redirects
  - 2.7 Distributed approaches
  - 2.8 Application-specific anti-spam methods
    - 2.8.1 RSS feed monitoring
    - 2.8.2 Response tokens
  - 2.9 Ajax
- 3 See also
- 4 References
- 5 External links

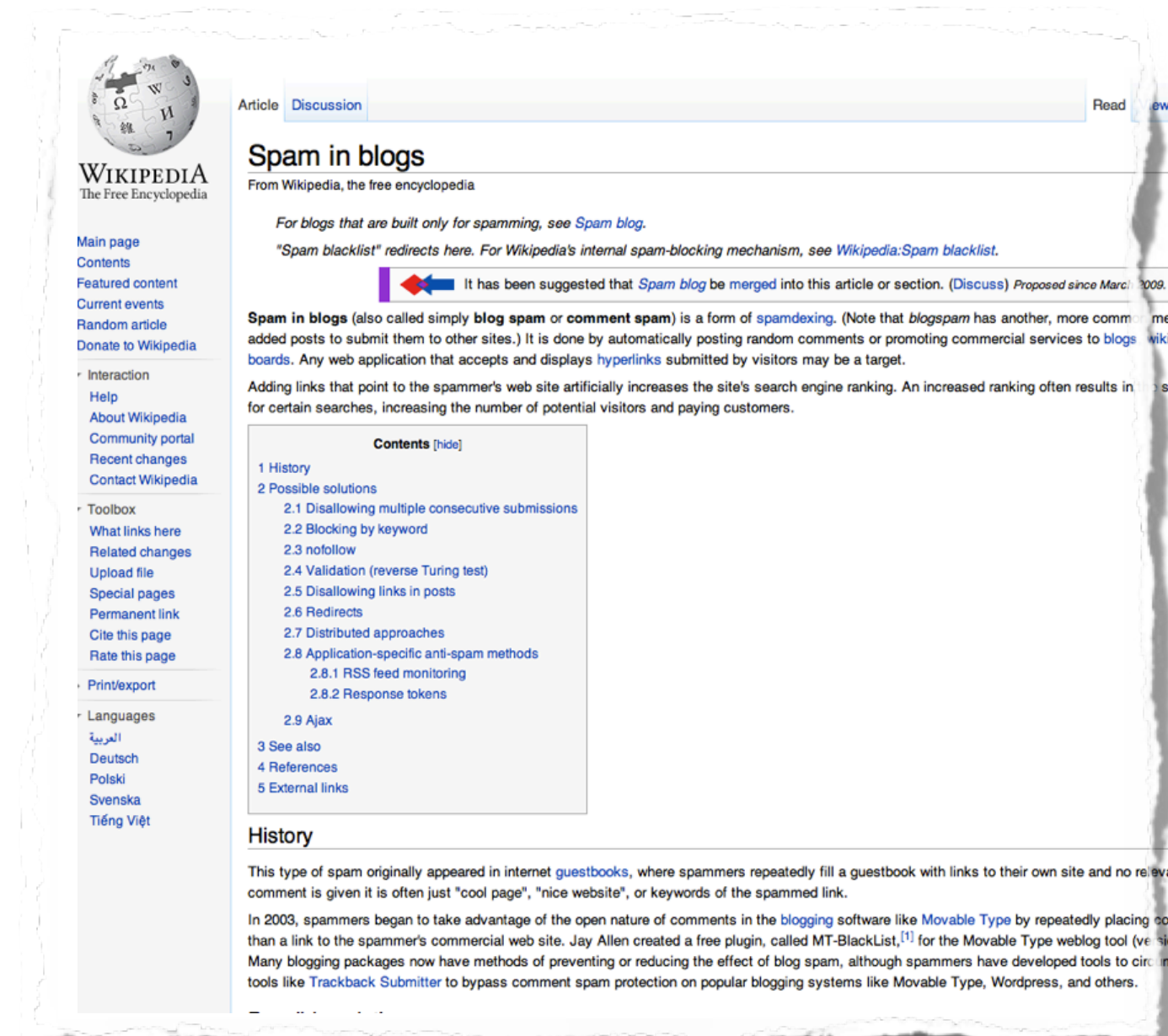
### History

This type of spam originally appeared in internet [guestbooks](#), where spammers repeatedly fill a guestbook with links to their own site and no relevant comment is given it is often just "cool page", "nice website", or keywords of the spammed link.

In 2003, spammers began to take advantage of the open nature of comments in the [blogging](#) software like [Movable Type](#) by repeatedly placing more than a link to the spammer's commercial web site. Jay Allen created a free plugin, called MT-BlackList,<sup>[1]</sup> for the Movable Type weblog tool (version 3.0.1). Many blogging packages now have methods of preventing or reducing the effect of blog spam, although spammers have developed tools to circumvent these tools like [Trackback Submitter](#) to bypass comment spam protection on popular blogging systems like Movable Type, Wordpress, and others.

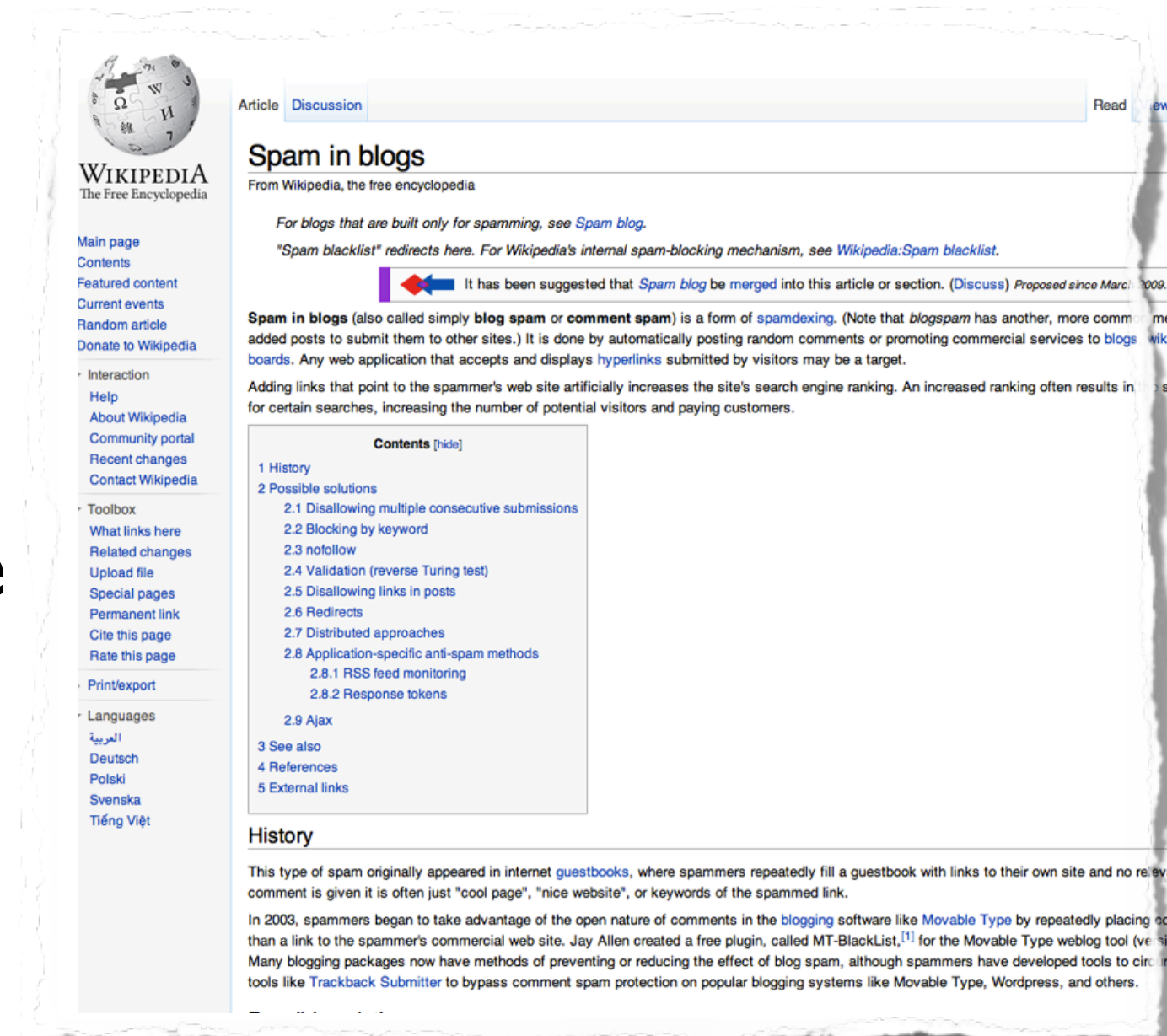
# Why the LinkBack problem is different ?

- A single spam can reach thousand of users



# Why the LinkBack problem is different ?

- A **single** spam can reach **thousand** of users
- LinkBack **notification** are **automated**

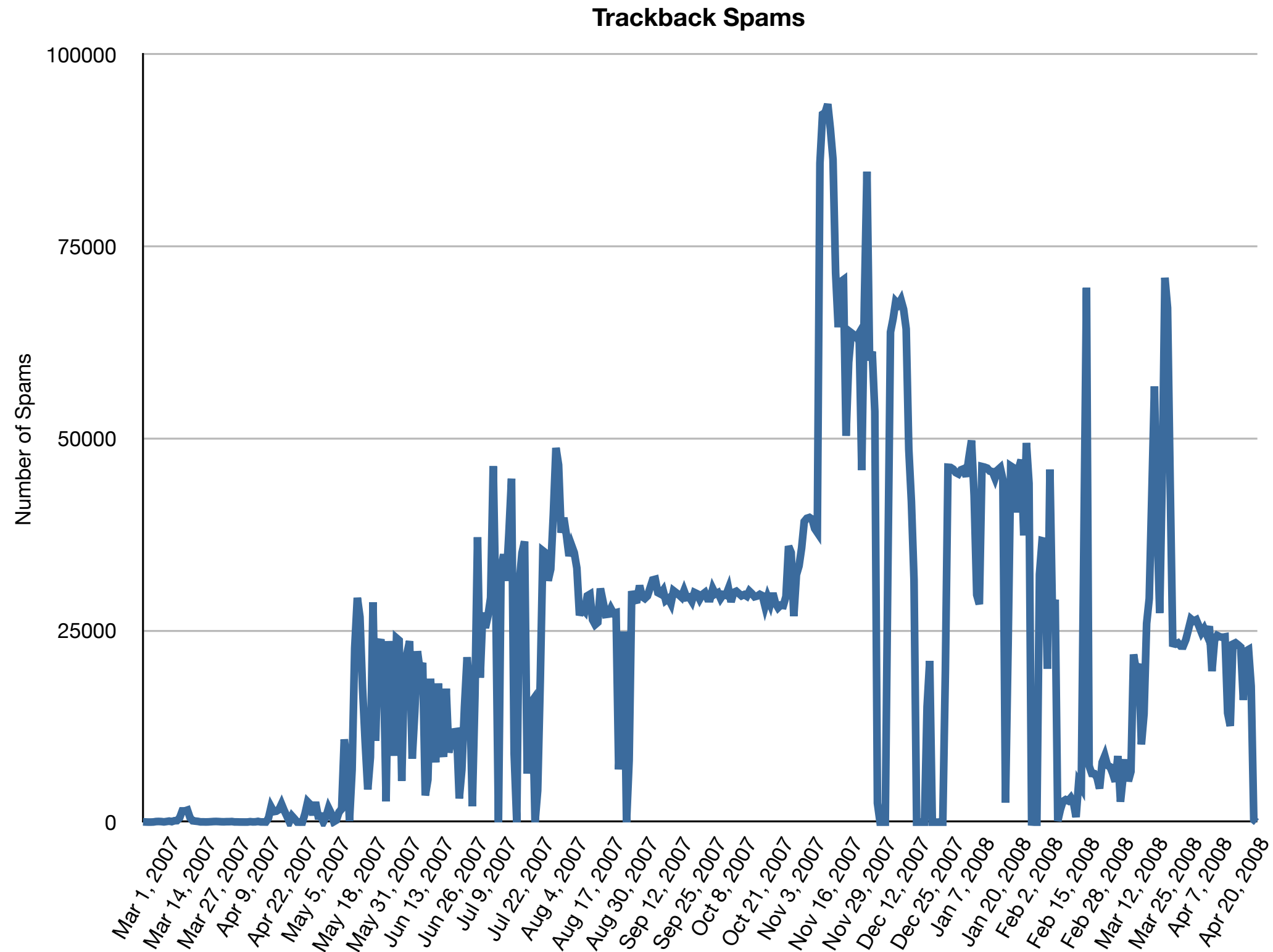






30 million LinkBack spam by day

# Spam Campaign example



# Spam Campaign example





# The end of all hope ?



## CODING HORROR

programming and human factors  
by Jeff Atwood

NEWER »

[Assertiveness for Software Developers](#)

OLDER

[Eric Lippert's Purple Crayon](#)

[Home](#)

[Browse All Posts](#)

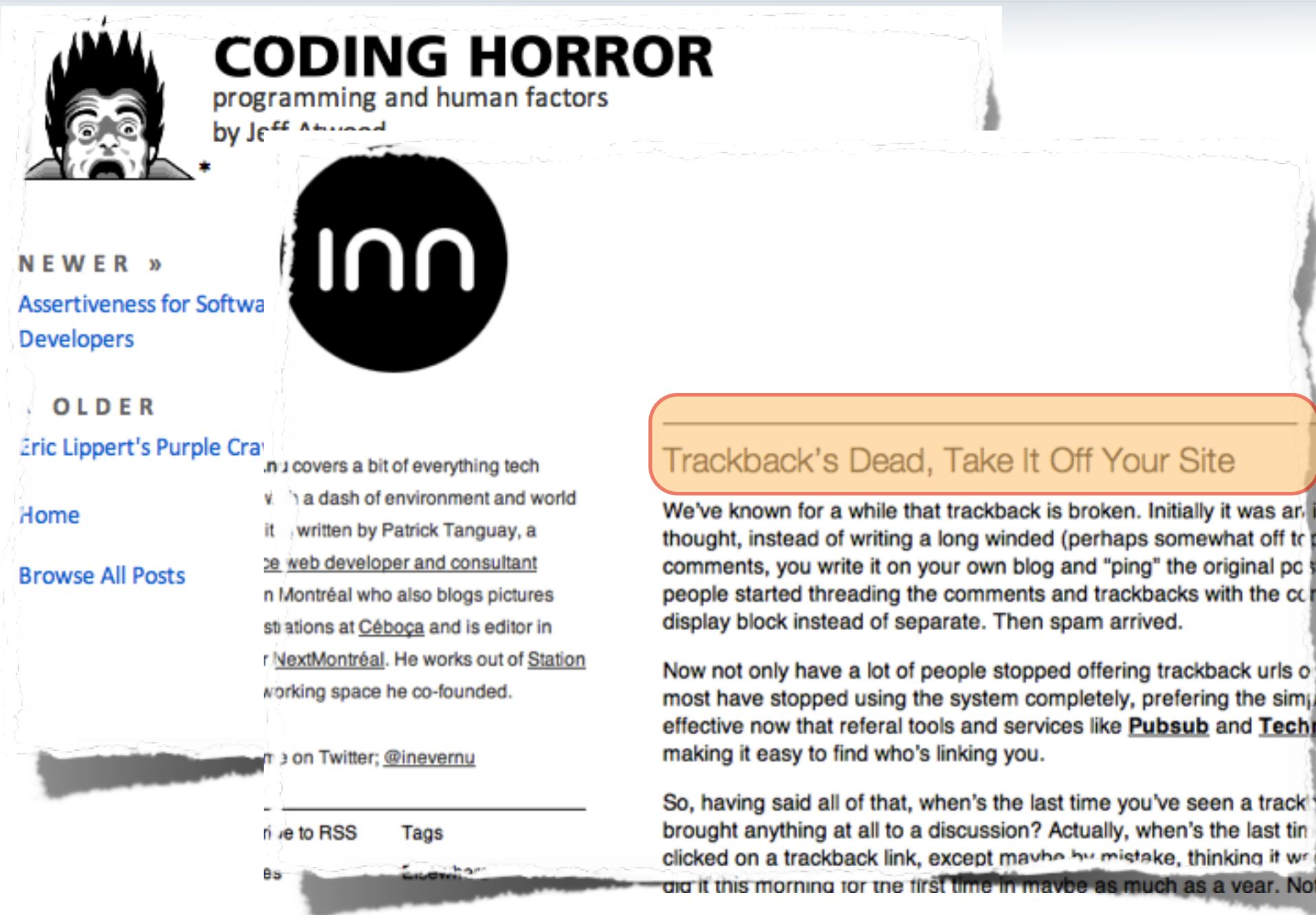
Dec 20, 2006

### The Day The Trackbacks Died

You might read a post on this blog and decide I feel this way. And, while you're at it, feel free to public dialog is why I believe so strongly in [code](#)

But sometimes a mere comment isn't enough. I write an entire post on *your* blog explaining, in you know that I, the target of your vitriol, have a little too intimate. Or, you could leave a comment

# The end of all hope ?




**CODING HORROR**  
programming and human factors  
by Jeff Atwood

**NEWER »**  
[Assertiveness for Software Developers](#)

**OLDER**  
[Eric Lippert's Purple Crayon](#)

[Home](#)  
[Browse All Posts](#)



... covers a bit of everything tech with a dash of environment and world it is written by Patrick Tanguay, a [web developer and consultant](#) in Montréal who also blogs pictures and musings at [Céboça](#) and is editor in chief of [NextMontréal](#). He works out of [Station](#) a working space he co-founded.

... on Twitter; [@inevernu](#)

[Subscribe to RSS](#)   [Tags](#)

[About](#)   [Privacy](#)   [Contact](#)


## Trackback's Dead, Take It Off Your Site

We've known for a while that trackback is broken. Initially it was an annoyance, instead of writing a long winded (perhaps somewhat off topic) comment, you write it on your own blog and "ping" the original post. Then people started threading the comments and trackbacks with the comment display block instead of separate. Then spam arrived.

Now not only have a lot of people stopped offering trackback urls on their blogs, most have stopped using the system completely, preferring the simpler and more effective now that referral tools and services like [Pubsub](#) and [Technorati](#) are making it easy to find who's linking you.


So, having said all of that, when's the last time you've seen a trackback brought anything at all to a discussion? Actually, when's the last time you've clicked on a trackback link, except maybe by mistake, thinking it was a comment? I did it this morning for the first time in maybe as much as a year. Not

# The end of all hope ?



## CODING HORROR

programming and human factors  
by Jeff Atwood



**NEWER »**  
[Assertiveness for Software Developers](#)

**OLDER**  
[Eric Lippert's Purple Crayon](#)

[Home](#)

[Browse All Posts](#)

inn covers a bit of everything tech with a dash of environment and world news. It is written by Patrick Tanguay, a web developer and consultant in Montréal who also blogs pictures of his cats at [Céboça](#) and is editor in chief of [NextMontréal](#). He works out of Station X, a working space he co-founded.

Follow me on Twitter; [@inevernu](#)

[Subscribe to RSS](#)   [Tags](#)

Be the first to like this.

## *the* BLOG HERALD

[news](#) • [editorial](#) • [features](#) • [interviews](#) • [guides](#) • [reviews](#) • [100 stories](#) • [archive](#) • [books](#) • [advertise](#)

### Have Trackbacks Become Too Spammy To Be Worthwhile?

You're at: [Home](#) • [Editorial](#) • Have Trackbacks Become Too Spammy To Be Worthwhile? • Filed as **Editorial, Features** on November 24, 2007 8:56 am

by **Tony Hung**



0



0



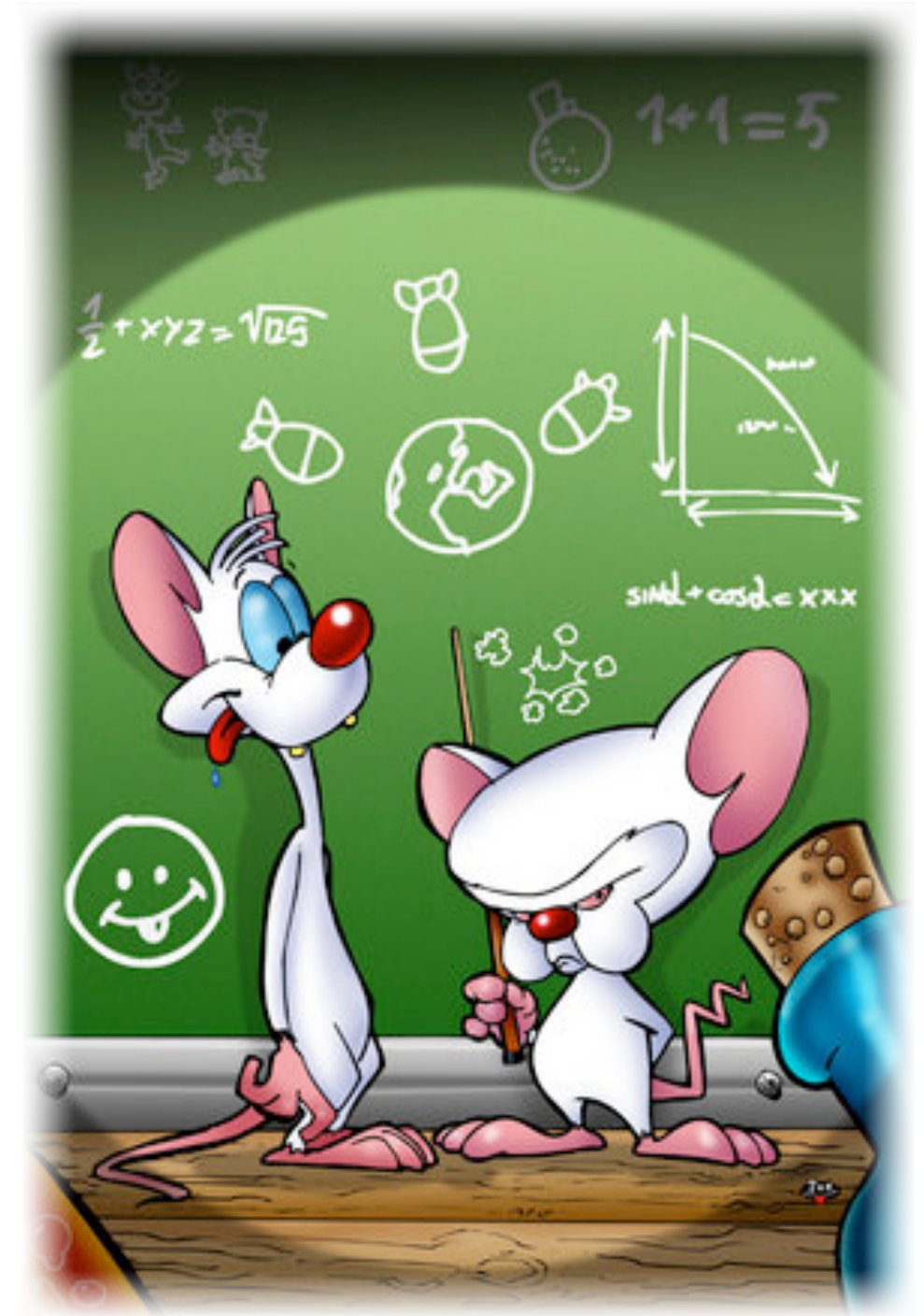
Be the first of your friends to like this.

You may have also discovered a surge in trackback spam recently as autoblogging software is being used by more and more spammers to reach out and cull RSS feeds. This phenomenon has led to many disabling trackbacks, or raising the "blacklist" level so high that you might never see some trackbacks again. Or, as some newer remotely-hosted commenting technologies like [IntenseDebate](#) and [Disqus](#) show, they simply do not show trackbacks because of the spam problem.



# Adversary Model

- Resourceful
- Efficient
- Knowledgeable
- Adaptive



# Threat

- Blog spoofing
- Cried Wolf attack
- Linkback tampering and replay
- Spam in breath or in deapth



# Current Linkback specifications

	RefBack	PingBack	TrackBack
Trigger Mechanism	Visit from the sender site	Code executed at posting time	Code executed at posting time
Notification medium	HTTP referer	XML-RPC call	<a href="#">HTTP POST</a>
Information sent	none	<ul style="list-style-type: none"> <li>- <i>S</i> post URL</li> <li>- <i>R</i> post URL</li> </ul>	<ul style="list-style-type: none"> <li>- <i>S</i> post URL</li> <li>- <i>S</i> site name</li> <li>- <i>S</i> post title</li> <li>- <i>S</i> post excerpt</li> </ul>
Auto-discovery mechanism	none	<b>LINK Tag</b>	Specially formatted information in the body





Threats are **not addressed** by current Linkback specifications

A background of red curtains with a dark grey horizontal band across the middle.

Introducing **Talkback** !



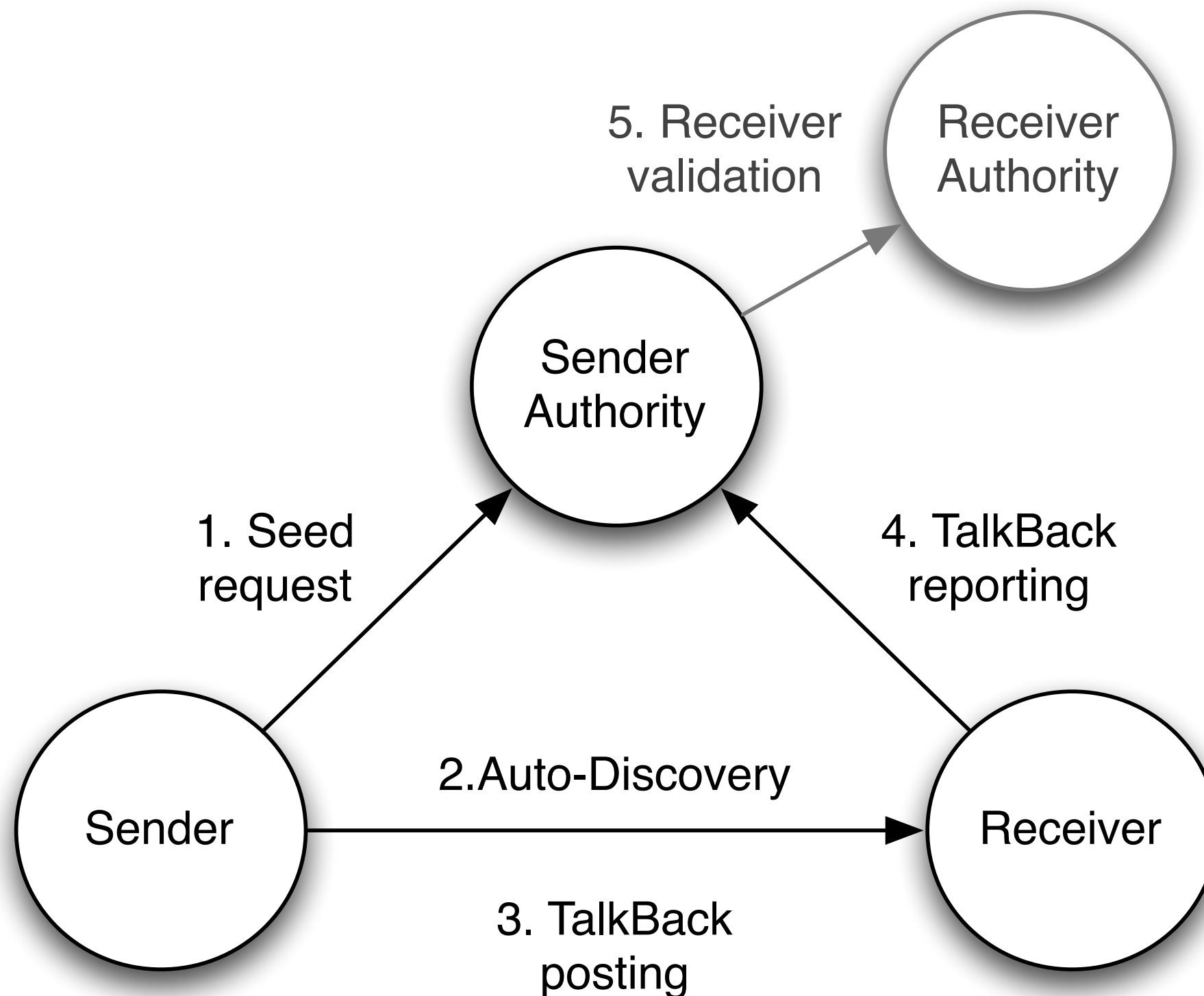
Use a **light-weight PKI** to fight  
Spam

# TalkBack Benefits

	RefBack	PingBack	TrackBack	TalkBack
Trigger Mechanism	Visit from the sender site	Code executed at posting time	Code executed at posting time	Code executed at posting time
Notification medium	HTTP referer	XML-RPC call	HTTP POST	HTTP POST
Information sent	none	<ul style="list-style-type: none"> <li>- <i>S</i> post URL</li> <li>- <i>R</i> post URL</li> </ul>	<ul style="list-style-type: none"> <li>- <i>S</i> post URL</li> <li>- <i>S</i> site name</li> <li>- <i>S</i> post title</li> <li>- <i>S</i> post excerpt</li> </ul>	<ul style="list-style-type: none"> <li>- <i>S</i> post URL</li> <li>- <i>S</i> site name</li> <li>- <i>S</i> post title</li> <li>- <i>S</i> post excerpt</li> <li>- Seed Token</li> <li>- <i>S</i> Public key</li> <li>- <i>R</i> Public key</li> <li>- Signature</li> </ul>
Auto-discovery mechanism	none	LINK Tag	Specially formatted information in the body	LINK Tag
<i>S</i> Authenticity	-	-	-	✓
<i>R</i> Authenticity	-	-	-	✓
Integrity	-	-	-	✓
Confidentiality	-	-	-	✓



# Talkback Overview



# Talkback Security Equation

---

# Talkback Security Equation

Limited number of blogs

---

# Talkback Security Equation

Limited number of blogs  
+ Limited number of LinkBack by blog

---



# Talkback Security Equation

$$\begin{array}{c} \text{Limited number of blogs} \\ + \text{ Limited number of LinkBack by blog} \\ \hline \text{Spam under-control} \end{array}$$

# Making blog registration costly

- Captcha
- Email verification
- Domain verification



# LinkBack benefits

- **Blog**
  - Authenticity
  - Whitelisting / blacklisting
- **Linkback**
  - Integrity
  - Non-repudiable
  - Confidentiality (optional)

# Reducing authority power

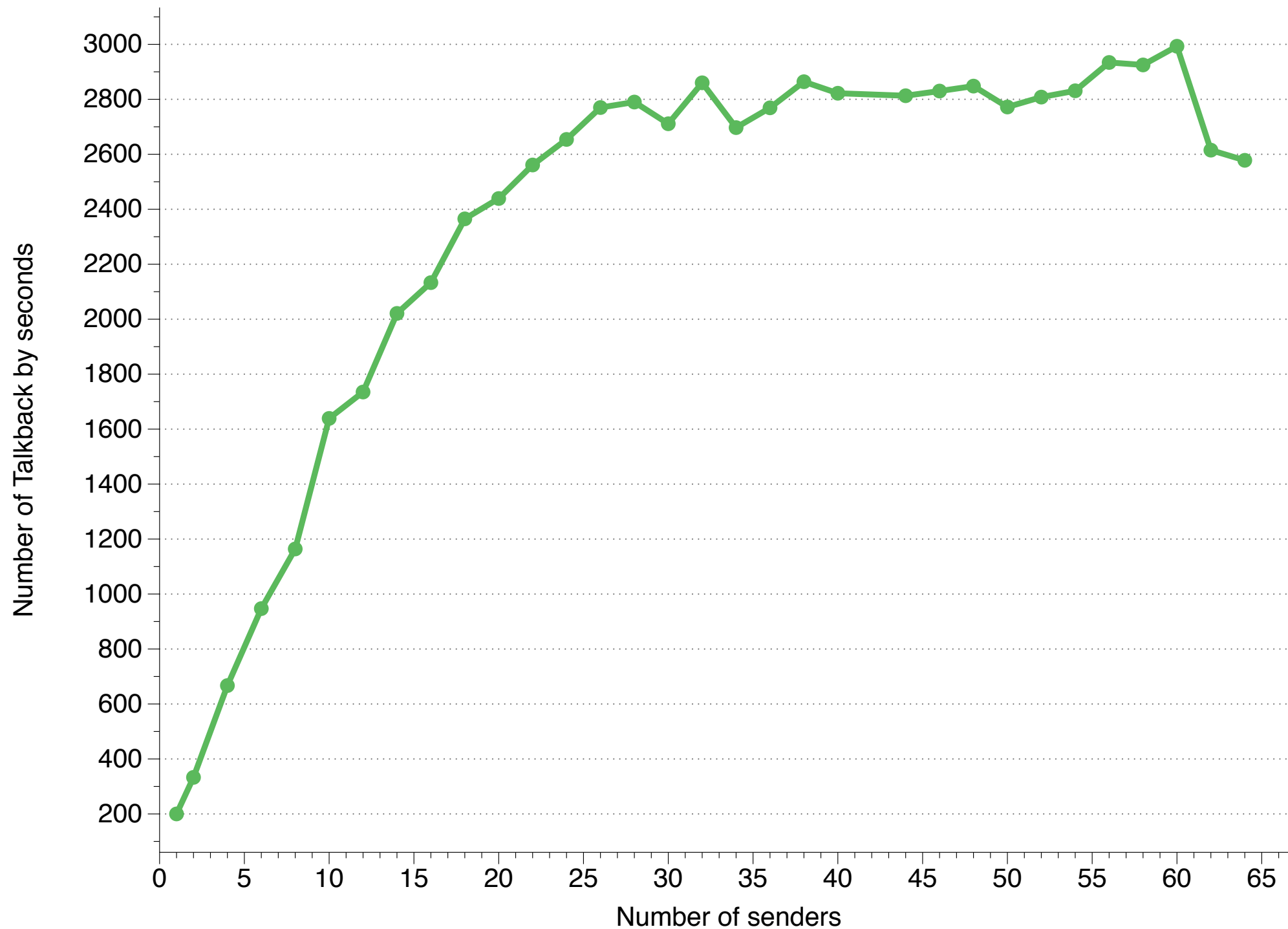
- Can't forge LinkBack: authorities have only blog public keys
- Respect privacy: authorities don't see LinkBack content in secure mode
- No single point of failure/control: protocol allows blog to choose the authority of their choice.

# Implementation

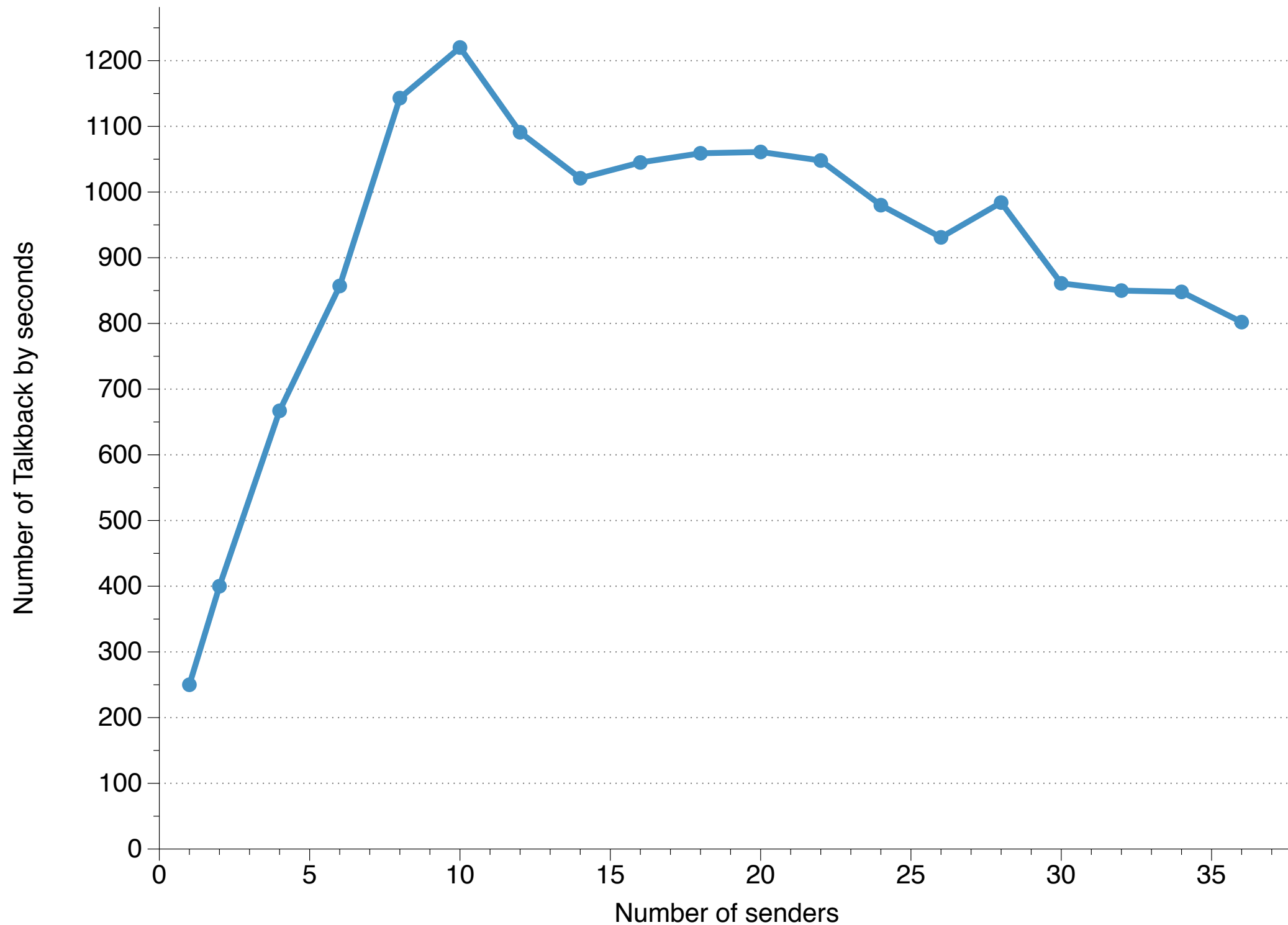
- Open-source library
- Wordpress plugin
- Authority : <https://talkback.stanford.edu>



# Authority benchmark



# Blog benchmark (Wordpress 3.1)





Thank you !

# Questions ?



Talkback

<http://ly.tl/talkback>

Follow-us on Twitter

[@elie](#), [@baptlste](#)