kaPoW Plugins

Protecting Web Applications Using Reputation-based Proof-of-Work

Tien Le, Akshay Dua, and Wu-chang Feng

Department of Computer Science
Portland State University

Portland State
UNIVERSITY

Spam Affects Many

- 4 million Facebook users
- receive spam from 600,000 hijacked accounts daily [2, 6]
- Email spam: 70.5% [1]

Spam Works

- In Jan 2010, 1.6 million Twitter users
- clicked 0.13% of all Twitter spam [2, 6]
- almost two orders of magnitude higher than email spam [3]

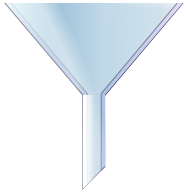
Spam is Costly

- Cost to businesses: $20.5 billion annually
- Projected to rise to $198 billion in four years [7].

Stopping Spam

- CAPTCHAs

- Spam filters



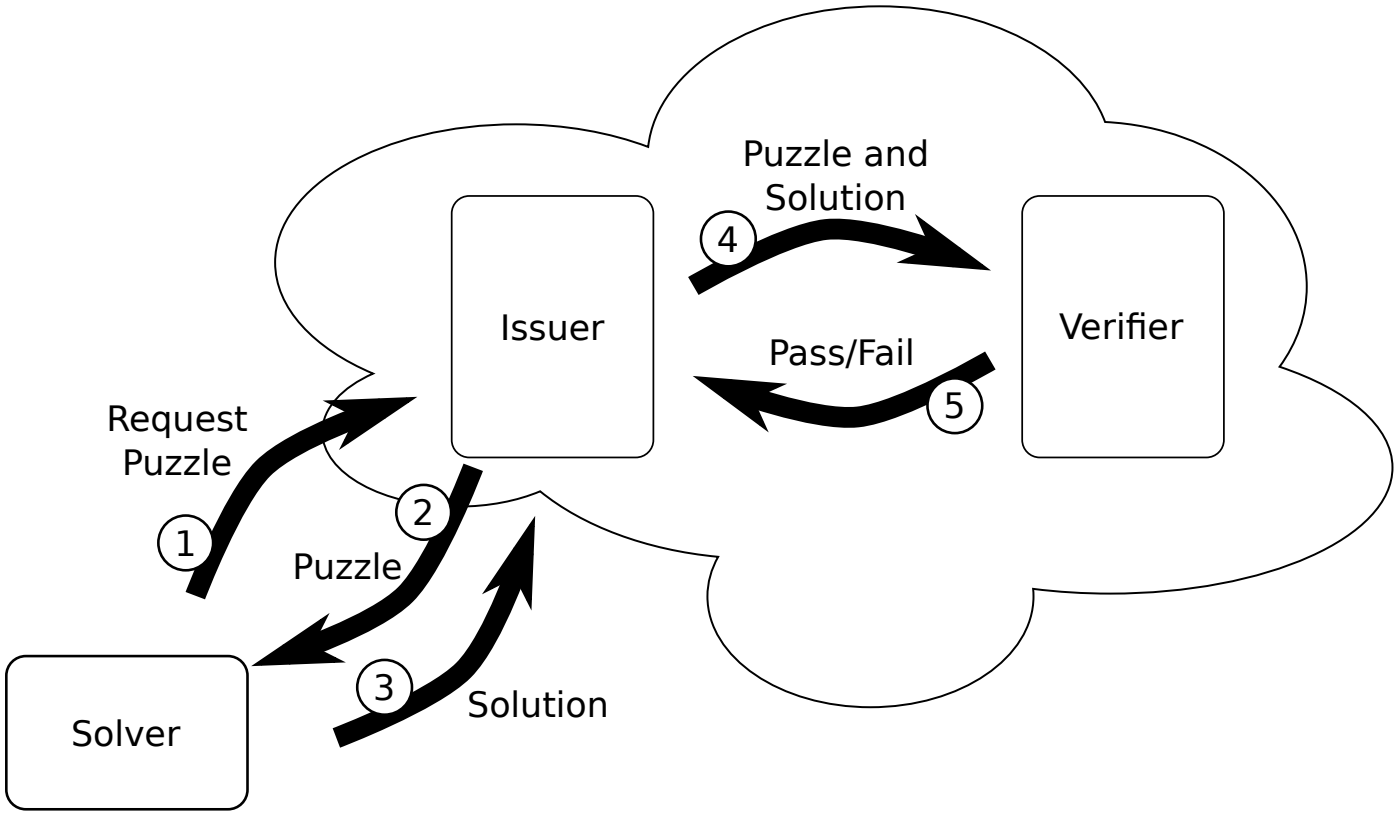
- User reports

We Can do More

Impose a cost per transaction (e.g. before posting a message)

"Proof-of-work" or *"client-puzzle"* approach

Proof-of-Work



Advantages

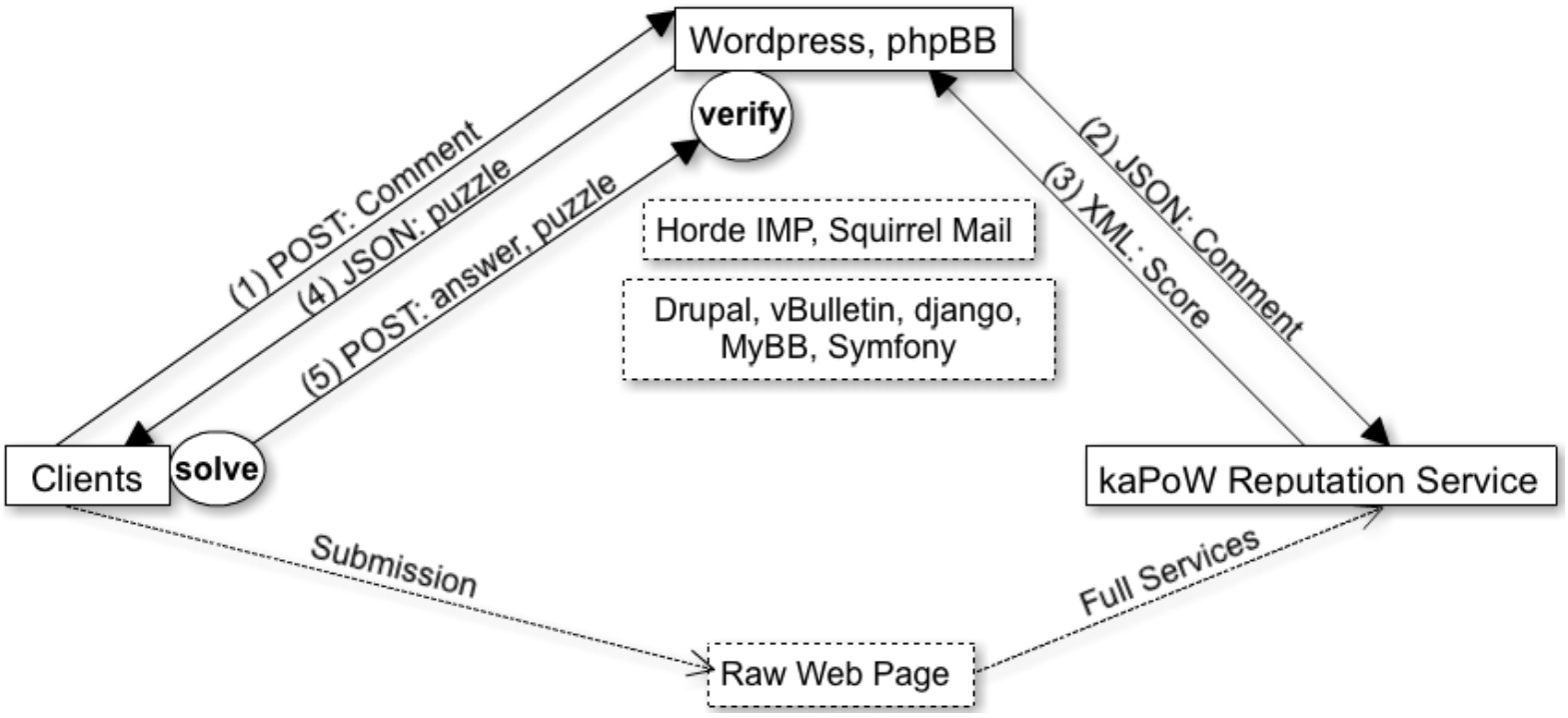
- *Slows* the rate of spam
- Can impose per-transaction cost
- Not a thumbs-up/down approach

Caveat

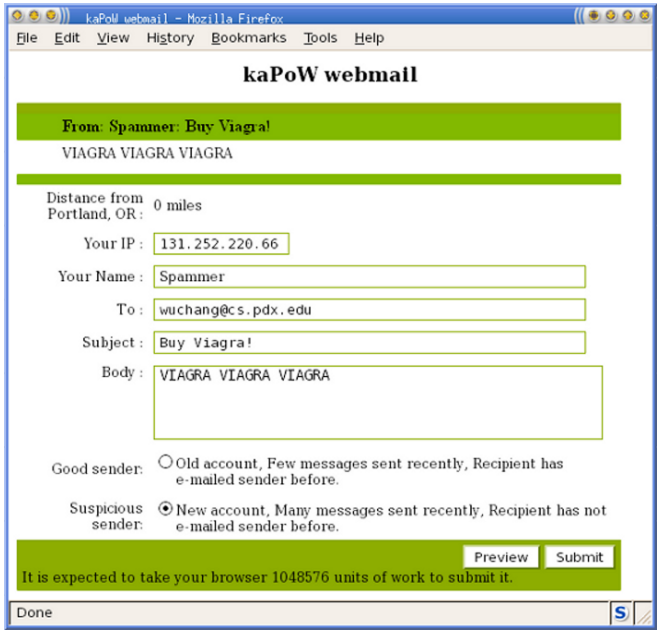
Successful only when *malicious* clients do *more* work [4, 5]

Need a *reputation system* and a way to set *puzzle difficulty*

kaPoW Plugins



Why kaPoW Plugins?



kaPoW Webmail not general enough

Needed a drop-in support for Proof-of-Work system

The "Work" in Proof-of-Work

kaPoW uses **Time-lock puzzles**

- Issuer chooses (a, t, n)
  - $n = p \cdot q$ for primes p, q
  - a: nonce $(1 < a < n)$
  - $t = \alpha \times T$: α squaring ops/sec for T secs
- Solver computes $A = a^{2^t}$ mod n
- Verifier's short-cut: $A' = a^{2^t \bmod \phi}$ mod n, where, $\phi = (p-1) \times (q-1)$
- Assert $A = A'$

Puzzle Difficulty

- Parameter $\mathbf{t}$ is puzzle difficulty [1]
- Difficulty based on *reputation score*

[1] in the time-lock puzzle

Reputation Score

- Weighted average of *local* and *global* scores
- score $= \sum_i b_i$, where $b_i \in \{0, 1\}$. b_i indicates...
  - *Is the IP address blacklisted?*
  - *Is the account new?*
  - etc.
- Difficulty $t = \alpha \times \text{score}^m$
  - $m =$ number of metrics (global and local)

kaPoW Plugins in Action

`http://kapow.cs.pdx.edu`

Portland State UNIVERSITY

Limitations

- Reputation metrics can be hard to determine
- Client message sent to kaPoW service
- All platforms are treated the same

References

Emailblog.eu.
Spam volume drops to historically low levels.
http://emailblog.eu/2011/12/16/spam-volume-drops-to-historical-low-levels/, Dec 2011.

Geoffrey A. Fowler, Shayndi Raice, Amir Efrati.
Facebook, Twitter battle 'social' spam.
http://www.theaustralian.com.au/business/wall-street-journal/facebook-twitter-battle-social-spam/story-fnay3ubk-1226237108998, Jan 2012.

C. Kanich, C. Kreibich, K. Levchenko, B. Enright, G. Voelker, V. Paxson, and S. Savage.
Spamalytics: An empirical analysis of spam marketing conversion.
In *Proceedings of the 15th ACM conference on Computer and communications security*, pages 3–14. ACM, 2008.

B. Laurie and R. Clayton.
'Proof-of-Work' Proves Not to Work'.
In *Workshop on Economics and Information Security*, May 2004.

D. Liu and L. Camp.
Proof of Work Can Work.
In *Workshop on Economics of Information Security*, June 2006.

Mark Risher.
Social Spam and Abuse — Annual Trend Review.
http://blog.impermium.com/2012/01/13/social-spam-and-abuse-the-year-in-review/, Jan 2012.

SPAM LAWS.
Spam Statistics and Facts.
http://www.spamlaws.com/spam-stats.html, 2011.