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
Spam Affects Many

- 4 million Facebook users
- receive spam from 600,000 hijacked accounts daily [2, 6]
- Email spam: 70.5% [1]
In Jan 2010, 1.6 million Twitter users clicked 0.13% of all Twitter spam [2, 6]. This is 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

1. Request Puzzle
2. Puzzle
3. Solution
4. Puzzle and Solution
5. Pass/Fail
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

Clients → solve → Wordpress, phpBB → verify → Horde IMP, Squirrel Mail → Drupal, vBulletin, django, MyBB, Symfony → kaPoW Reputation Service

Submission → Raw Web Page → Full Services

1. POST: Comment
2. JSON: Comment
3. XML: Score
4. JSON: puzzle
5. POST: answer_puzzle
Why kaPoW Plugins?

kaPoW Webmail not general enough

*Needed a drop-in support for Proof-of-Work system*
kaPoW uses **Time-lock puzzles**

- **Issuer chooses** $(a, t, n)$
  - $n = p \cdot q$ for primes $p, q$
  - $a$: nonce $(1 < a < n)$
  - $t = \alpha \times T$: $\alpha$ squaring ops/sec for $T$ secs

- **Solver computes** $A = a^{2^t} \mod n$

- **Verifier’s short-cut:** $A' = a^{2^t \mod \phi \mod n}$, where, $\phi = (p - 1) \times (q - 1)$

- **Assert** $A = A'$
Puzzle Difficulty

- Parameter $t$ is puzzle difficulty
- Difficulty based on *reputation score*

---

$^1$ in the time-lock puzzle
Reputation Score

- Weighted average of local and global scores
  
  \[ \text{score} = \sum_i b_i, \text{ where } b_i \in \{0, 1\}. \text{ } b_i \text{ indicates...} \]
  
  - Is the IP address blacklisted?
  - Is the account new?
  - etc.

- Difficulty \( t = \alpha \times \text{score}^m \)
  
  - \( m = \text{number of metrics (global and local)} \)
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.

Geoffrey A. Fowler, Shayndi Raice, Amir Efrati.
Facebook, Twitter battle 'social' spam.

Spamalytics: An empirical analysis of spam marketing conversion.

B. Laurie and R. Clayton.
‘Proof-of-Work’ Proves Not to Work’.

D. Liu and L. Camp.
Proof of Work Can Work.

Mark Risher.
Social Spam and Abuse — Annual Trend Review.

SPAM LAWS.
Spam Statistics and Facts.