RBLs evolution

from IP reputation to the next generation

Giovanni Bechis, CTO @ Pccc

PCCC

Who am I ?

CTO @ Pccc

Apache SpamAssassin v.p.

RBLs to detect email abuse

Real time detection of abused email red flags

Dns based solution, fast and easy to scale

Ip based RBL

Spammers are using open relays or stolen credentials

Checks ip addresses in received headers

Detects mail servers abuse

Domain based RBL

- Spammers are using a pool of shared ip addresses
- Checks domains in uris and email addresses

Email address RBL

Attackers rotate email addresses for scams

Blocklists like MSBL track abusive email addresses (md5 hashes stored in dns zones)

Bitcoin wallets RBL

- Ransomware operators collect payments using Bitcoin wallets
- Wallet blocklists flag addresses linked to illegal activities

ESP abused accounts

- Spammers are abusing ESP accounts to send emails using someone else‘s infrastructure

- ESP blocklists lists abused ESP accounts

Phone numbers RBL

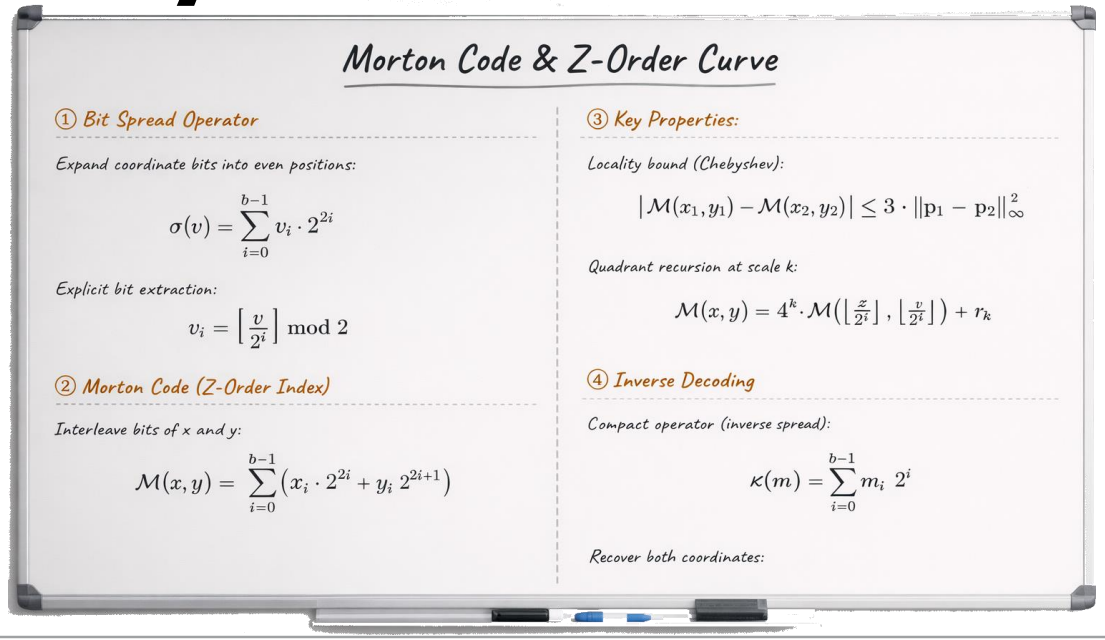
Scammers use disposable phone numbers for fraud

Phone number blocklists allow systems to detect spam frauds

Fuzzyhash RBL

- Spammers use templates to send emails to victims
- Software like Vipul Razor, Pyzor or DCC detect emails similar to known spam messages

Fuzzyhash RBL



Morton Code & Z-Order Curve

① Bit Spread Operator

Expand coordinate bits into even positions:

$$\sigma(v) = \sum_{i=0}^{b-1} v_i \cdot 2^{2i}$$

Explicit bit extraction:

$$v_i = \left\lfloor \frac{v}{2^i} \right\rfloor \bmod 2$$

② Morton Code (Z-Order Index)

Interleave bits of x and y:

$$\mathcal{M}(x,y) = \sum_{i=0}^{b-1} \left(x_i \cdot 2^{2i} + y_i \, 2^{2i+1} \right)$$

③ Key Properties:

Locality bound (Chebyshev):

$$|\mathcal{M}(x_1,y_1) - \mathcal{M}(x_2,y_2)| \leq 3 \cdot \|\mathrm{p}_1 - \mathrm{p}_2\|_\infty^2$$

Quadrant recursion at scale k:

$$\mathcal{M}(x,y) = 4^k \cdot \mathcal{M}\left(\left\lfloor \frac{x}{2^i} \right\rfloor , \left\lfloor \frac{v}{2^i} \right\rfloor \right) + r_k$$

④ Inverse Decoding

Compact operator (inverse spread):

$$\kappa(m) = \sum_{i=0}^{b-1} m_i \; 2^i$$

Recover both coordinates:

Fuzzyhash RBL

The rendered message body is passed through ZOrder's minhash, which normalizes tokens and produces a 32-bit signature.

Fuzzyhash RBL

The 32-bit signature is split into 4 bands of 8 bits.
Each band becomes a DNS label (fz2<band><value>.<domain>), yielding exactly 4 TXT queries.
Two texts sharing a band means all 8 bits matched in that slice, this acts as a locality-sensitive hash to cheaply find candidate matches.

Fuzzyhash RBL

Each TXT response carries the full 32-bit digest of an indexed string.

The algorithm computes the hamming distance between the query digest and stored digest.

If similarity >= *sim_threshold*, the email is similar to a spam message.

Unrelated texts score ~50%.

Fuzzyhash RBL

False positive rate

Similarity threshold 100%: ~9×10^{-10}

Similarity threshold 95%: ~2.3×10^{-8}

Fuzzyhash RBL

Generally available from May 2026

in production since the beginning of the year


Thanks!
Questions?
Q&A?