IBM Research

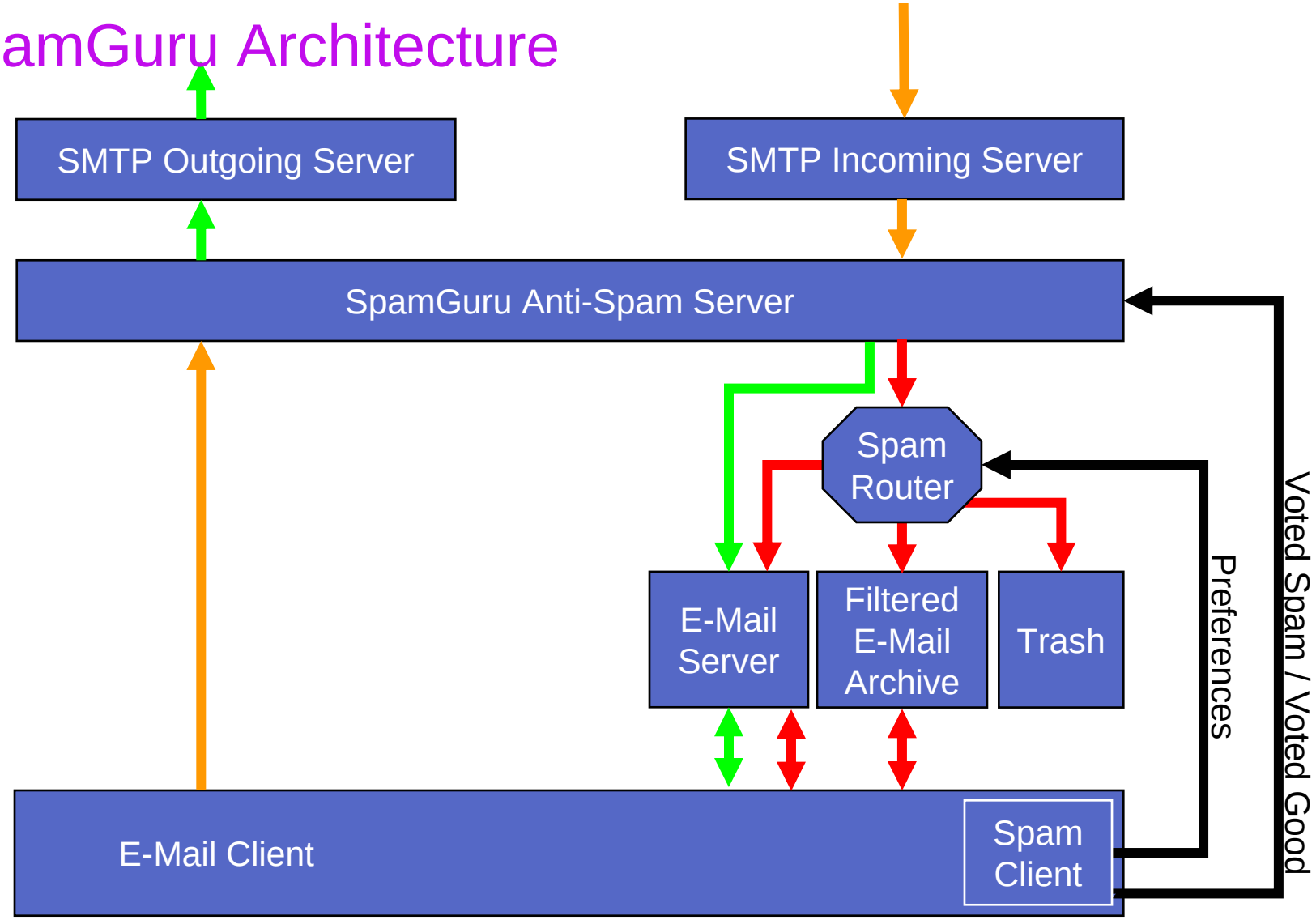
IBM SpamGuru on the TREC 2005 Spam Track

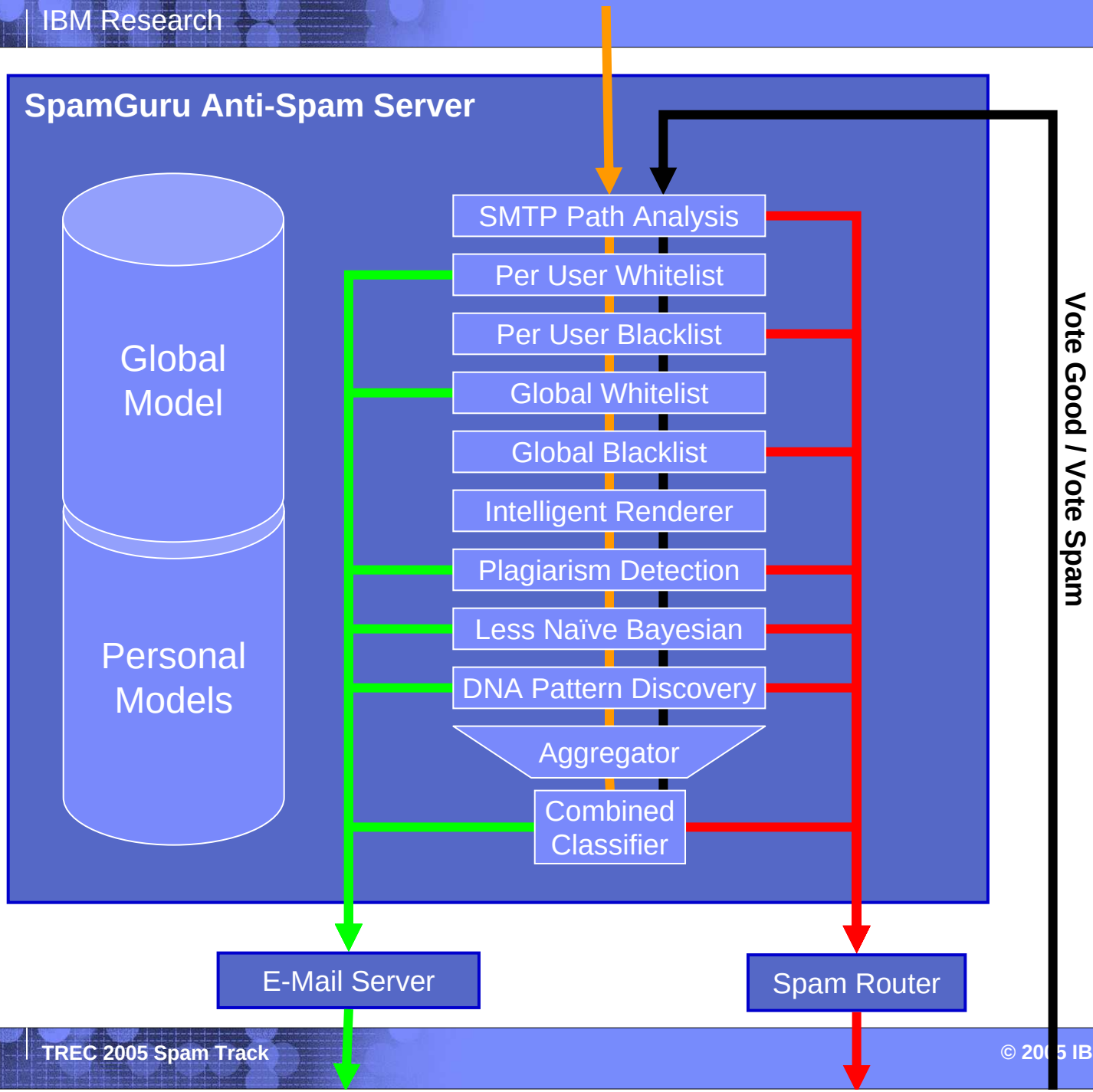
Richard Segal

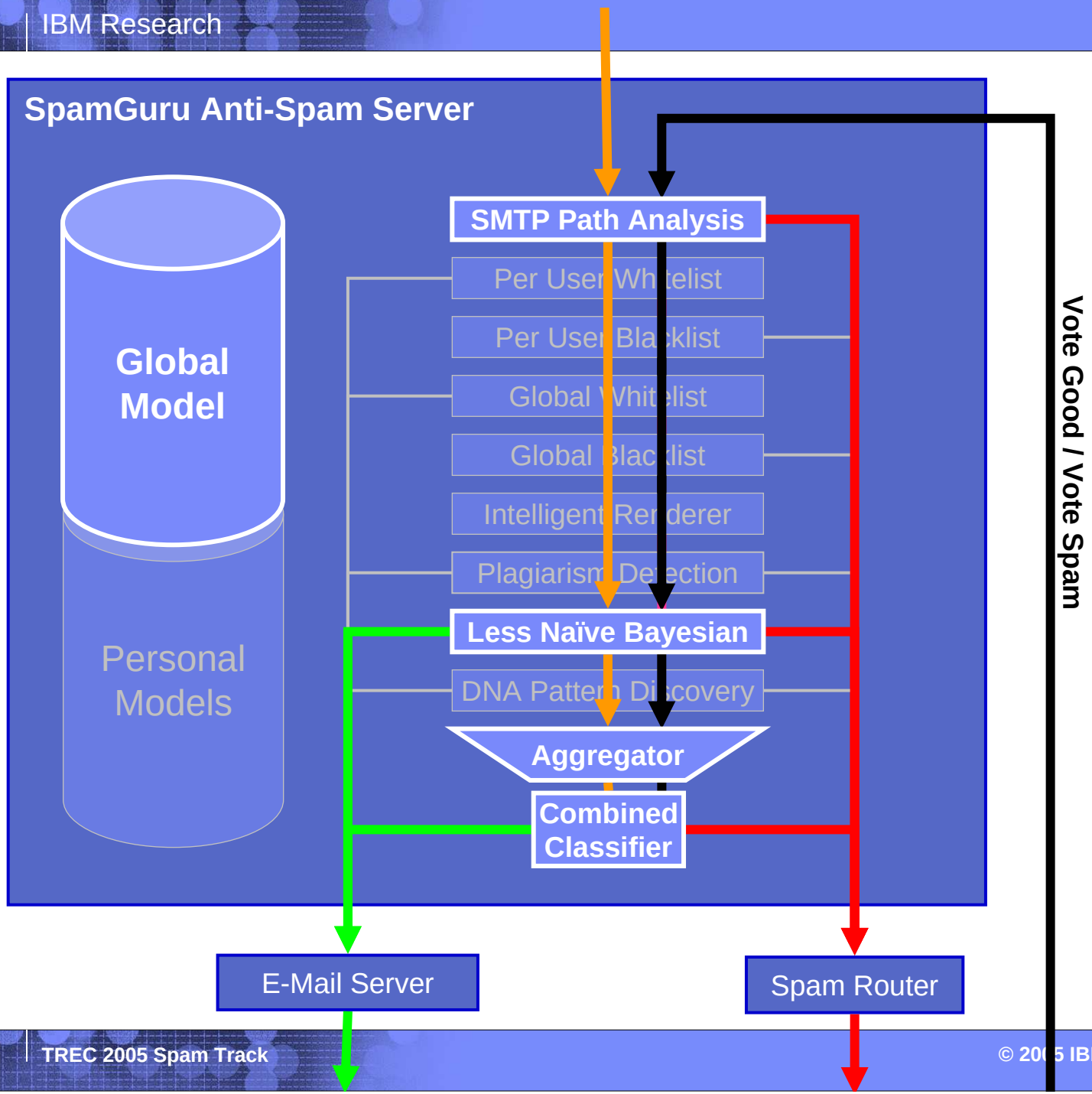
Jason Crawford, Jeff Kephart, Barry Leiba,
V.T. Rajan, Mark Wegman

November 18, 2005

SpamGuru Architecture







Classifier Aggregation

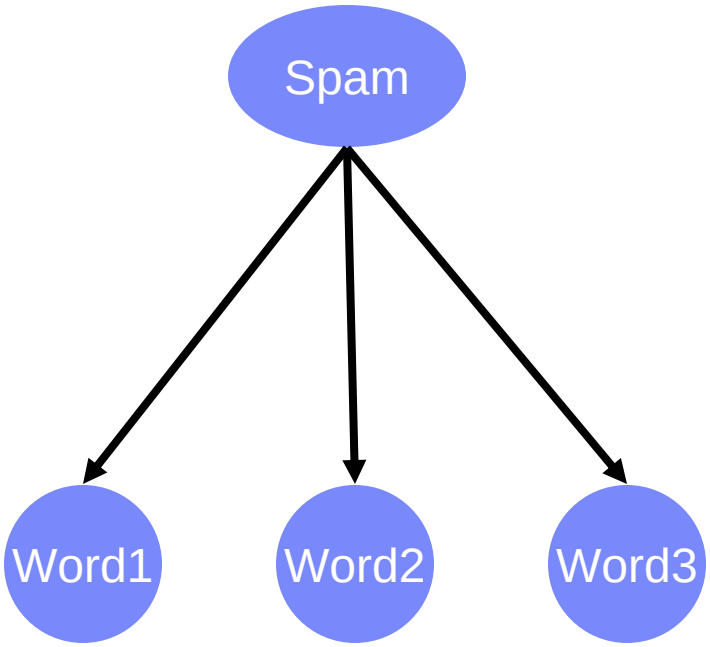
- Combine results of multiple classifiers to predict spam
- Catch more spam with less false positives
  - Emphasize each algorithms strengths.
  - De-emphasize each algorithms weaknesses.
- Harder to attack
  - Must simultaneously break through multiple algorithms.
- Adapts to changes in classifier effectiveness

Aggregation by Optimized Linear Weights

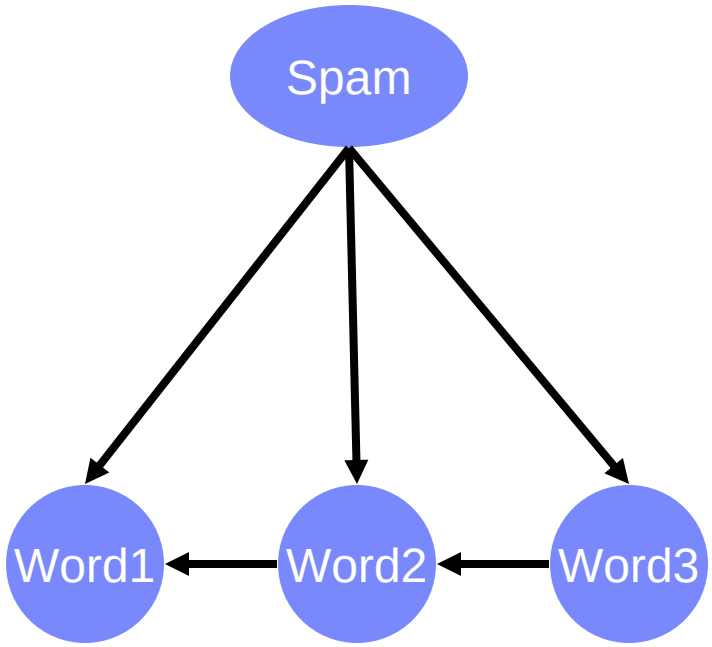
$$Score(x) = b + \prod \left(W_i \times Score_i(x) \right)$$

- Find optimal values for W_i and b using a Nelder-Mead non-linear optimizer.
- Re-optimize values every 10,000 examples.

Naïve Bayesian

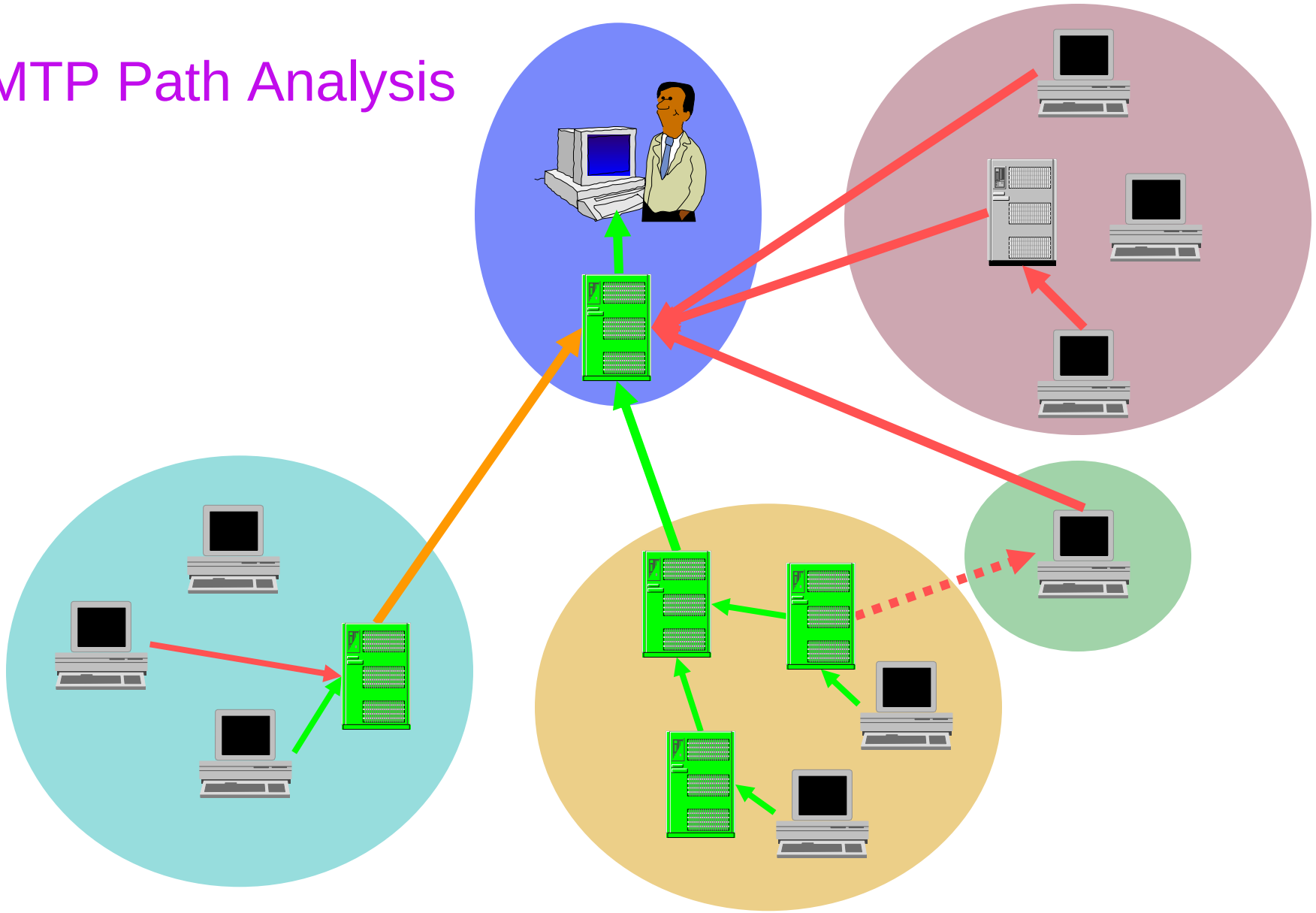


Less Naïve Bayesian



- Approximate calculation using a greedy update rule.
- Very little additional computational cost.
- Big performance boost.

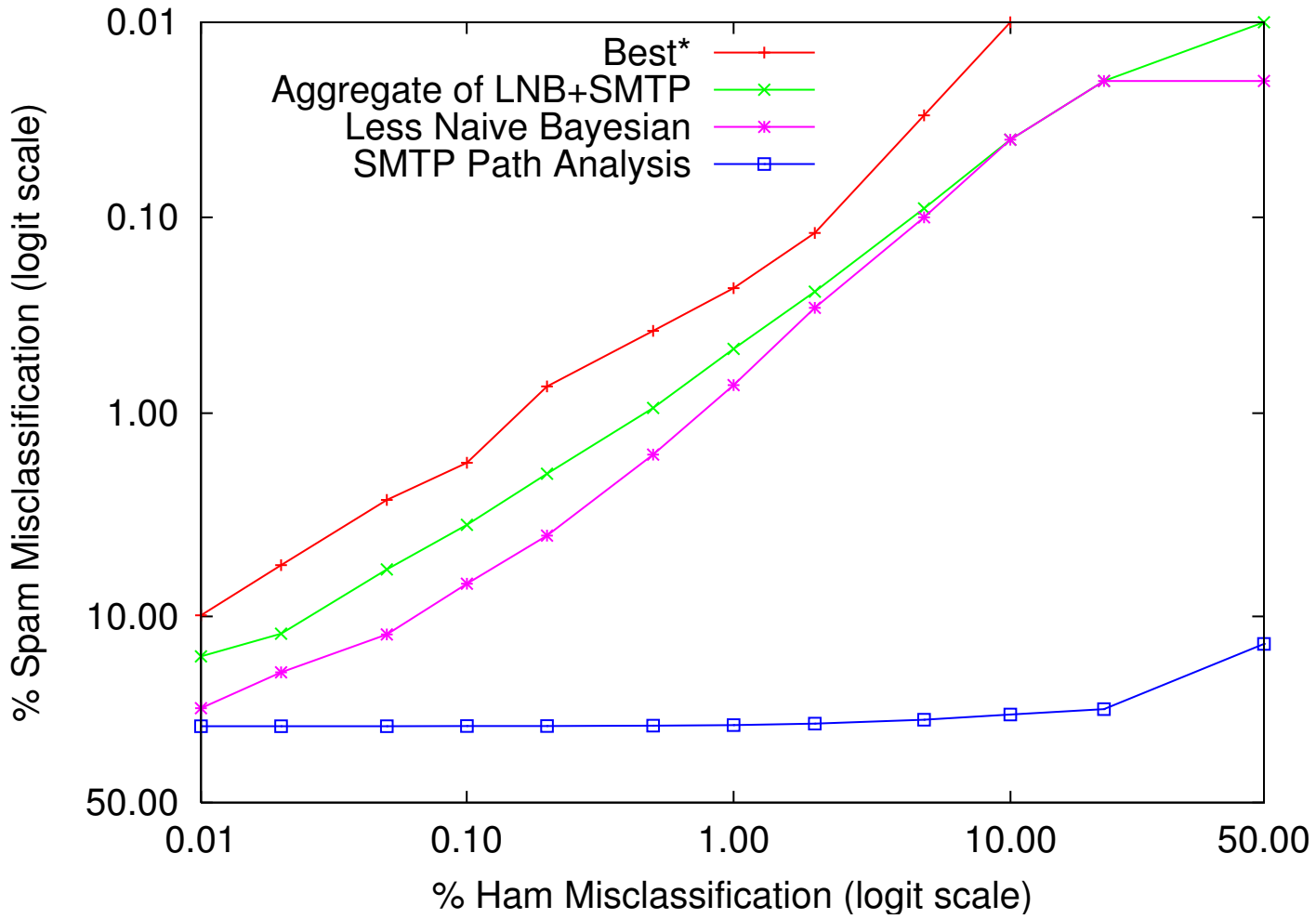
SMTP Path Analysis



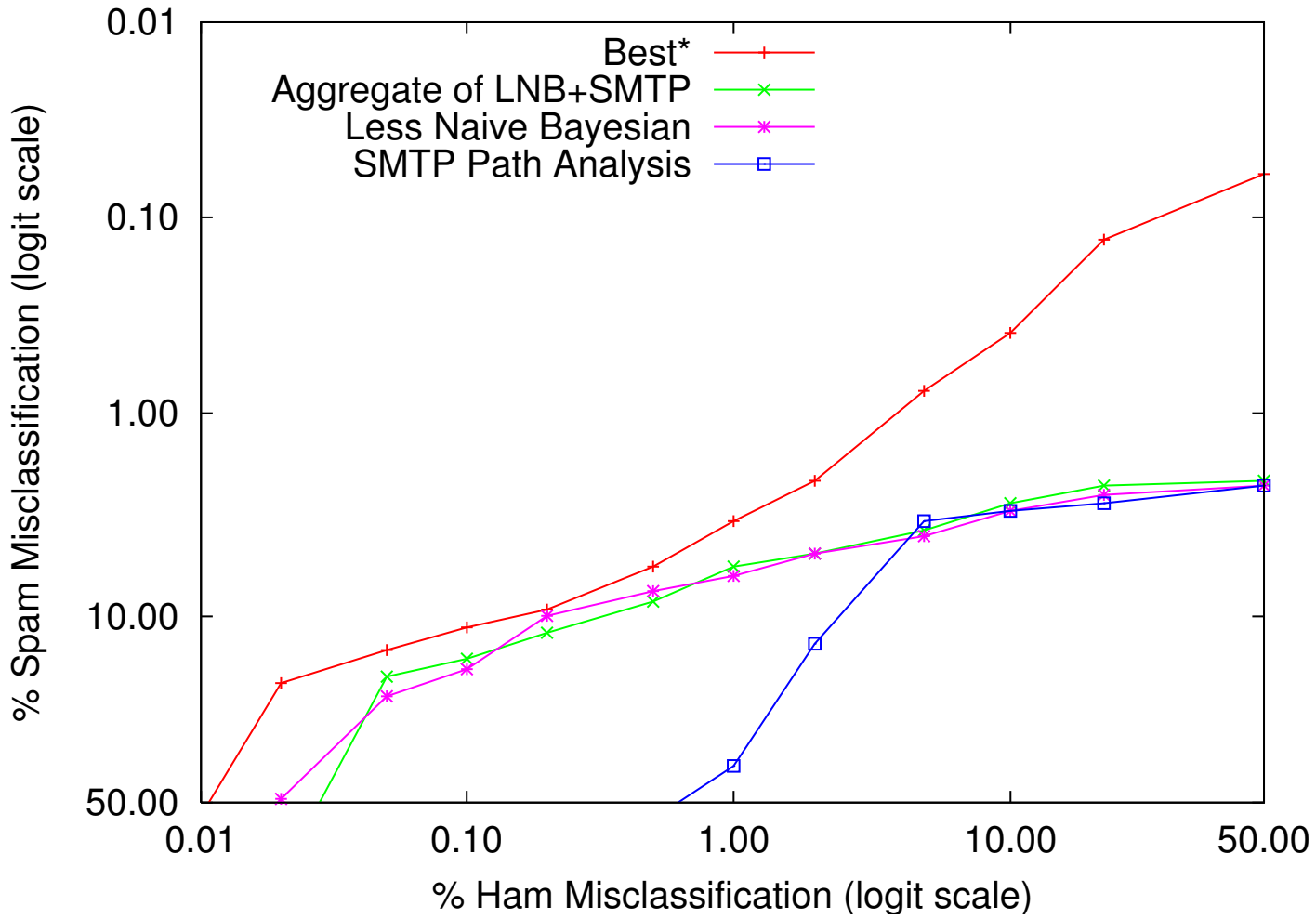
Submission Details

- General
  - Pre-trained on 20,000 labeled messages.
  - Corpus created from honeypots and user voting records.
- Text Processing
  - Decoded MIME but left attachments in place.
  - Word-based tokens. No stop words, no feature selection.
  - Special handling of URLs, e-mail addresses, etc.
- TREC Required Bug (minor)
  - Classified all messages longer than 100K as good.

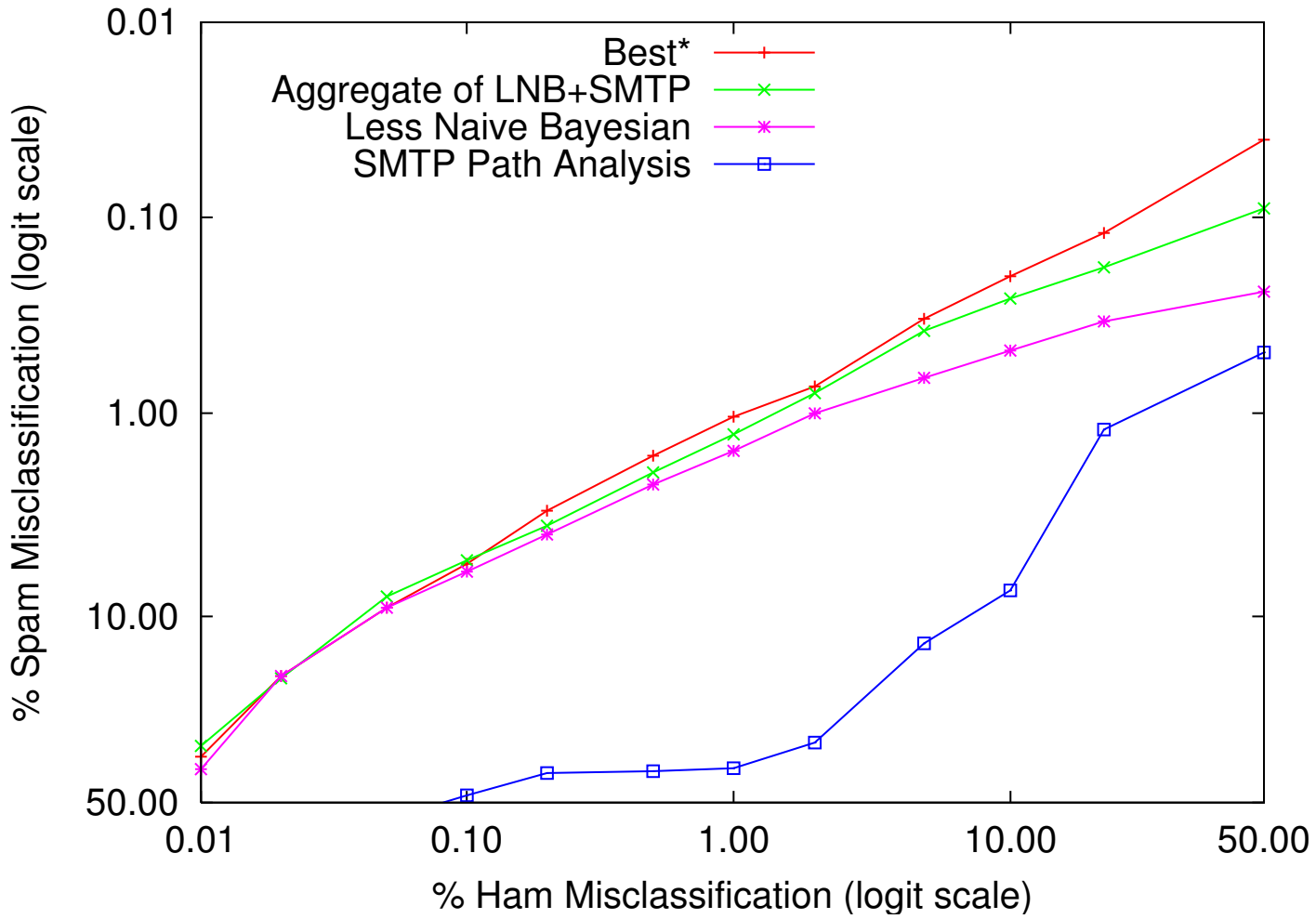
“Full” Dataset



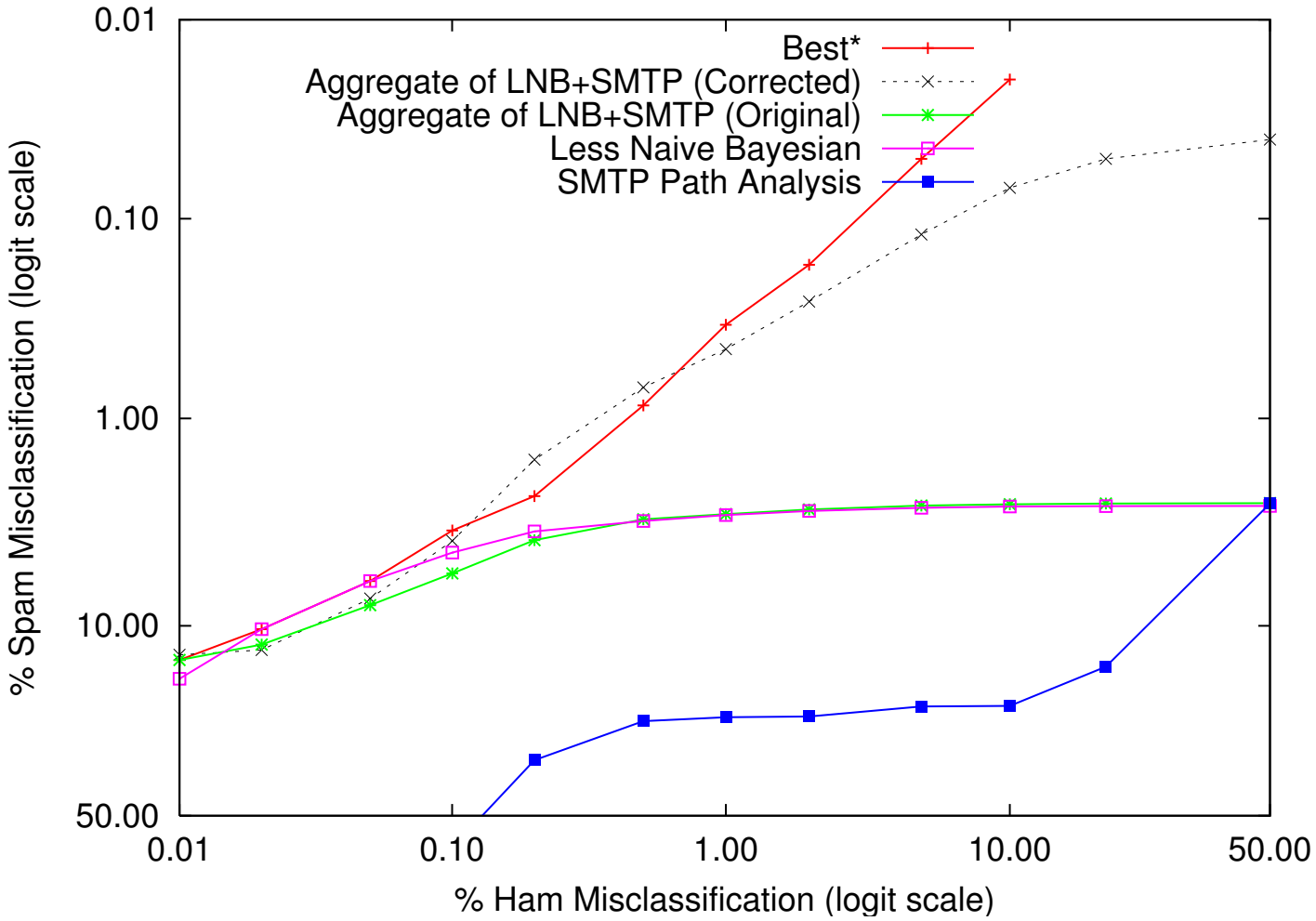
“SB” Dataset



“TM” Dataset



“Mr. X” Dataset



Summary

- Classifier aggregation using non-linear optimization.
- Less-naïve Bayesian performs well.
- SMTP path analysis is not very good in isolation, but combines well with Less-naïve Bayesian.

http://www.research.ibm.com/spam