Anti-Spam Mythology: Testing Trumps PR

Joel Snyder
jms@opus1.com
Opus One

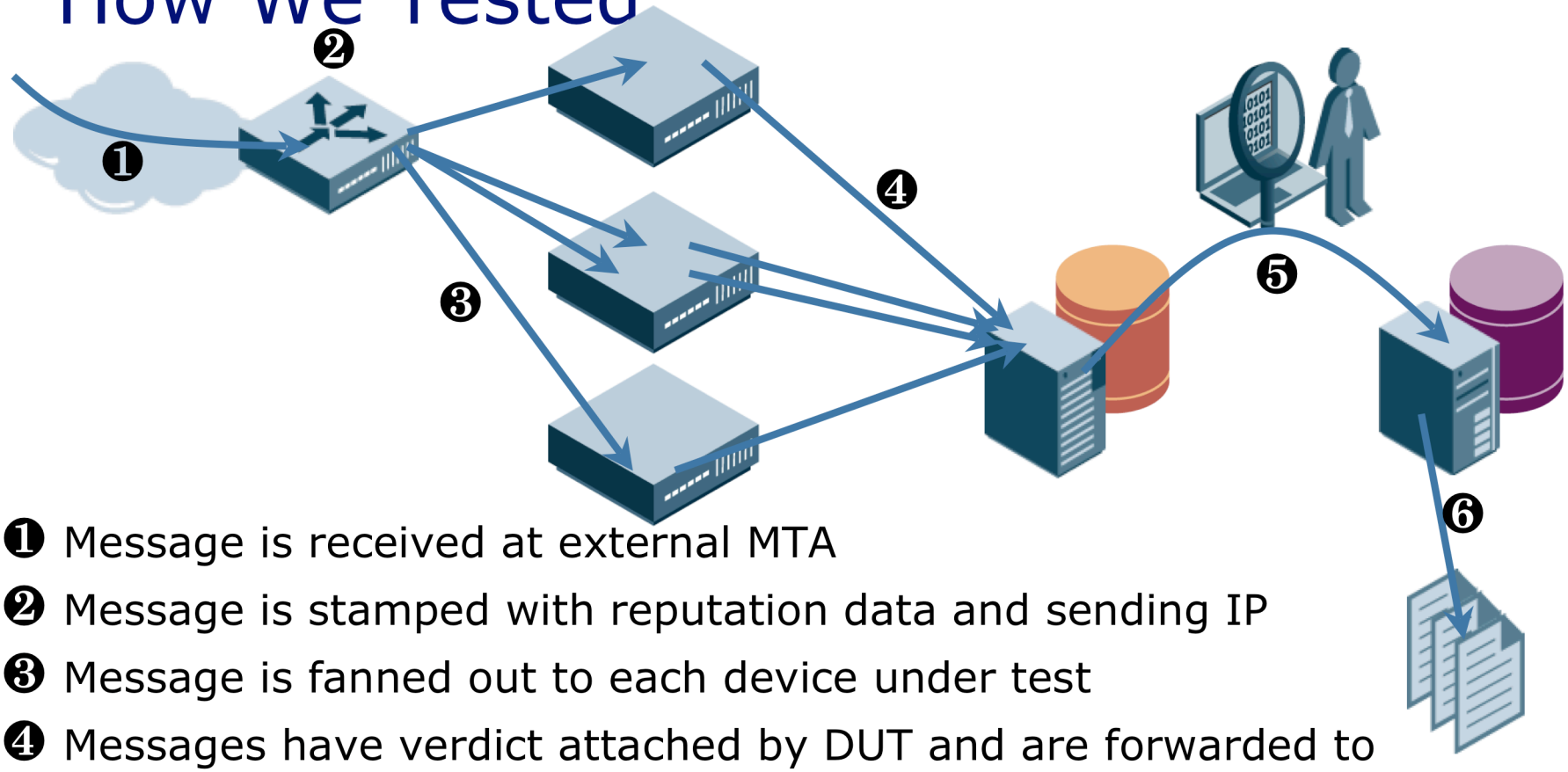
Agenda

- **What's all this testing nonsense anyway?**
- **Top Myths in Running Enterprise Anti-Spam**

Where Did All This Come From?

- **Network World anti-spam tests in Aug/2003 and Dec/2004**
- **First (and only) public test using real message streams**

"How We Tested"



❶ Message is received at external MTA

❷ Message is stamped with reputation data and sending IP

❸ Message is fanned out to each device under test

❹ Messages have verdict attached by DUT and are forwarded to mailbox server

❺ Master set is combed through by human to classify as spam/not-spam/don't-know; FPs are classified as 1 to 5 by "badness"

❻ Many SQL queries later, a report is put together (partially by hand)

"When We Tested"

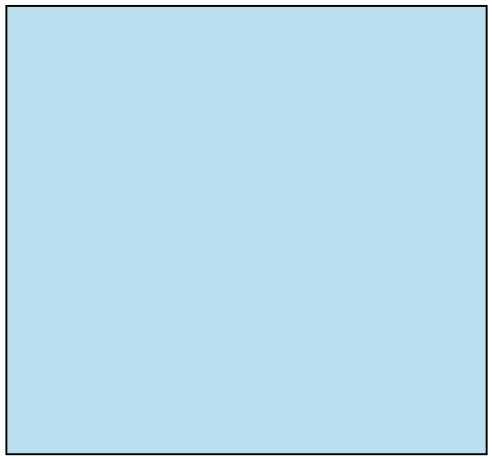
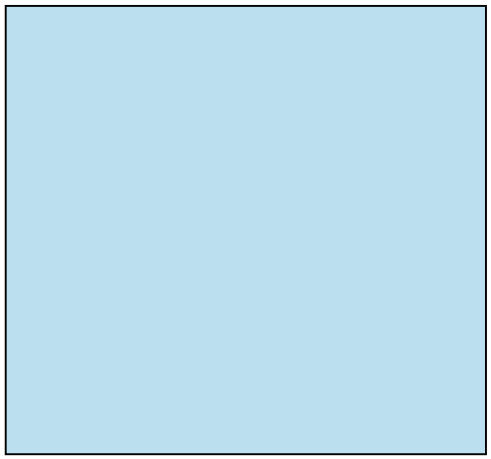
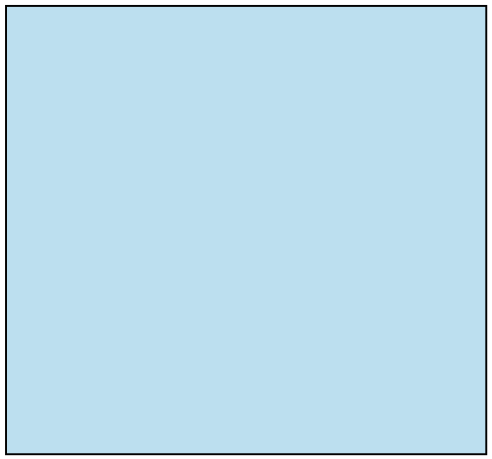
- **Three tests in First Half of 2005**
- **Starting Second Half of 2005, approximately quarterly testing (17 runs as of Feb/2008)**
- **Each test looks at between 15 and 45 scenarios**
- **…and a continuing test plan through 2008**

"Who We Tested"

- **Each run had different participants, but the data include runs from…**
- **Abaca, Barracuda, Brightmail, Ciphertrust, Cloudmark, Commtouch, Exchange 2007, Frontbridge, GFI, Heluna, Ironport, MailShell, Postini, Proofpoint, SonicWALL, Sophos, SpamStopsHere, StarEngine, Threatwall, Trend Micro, Tumbleweed, Untangle, Watchguard and 2 others that don't have names**

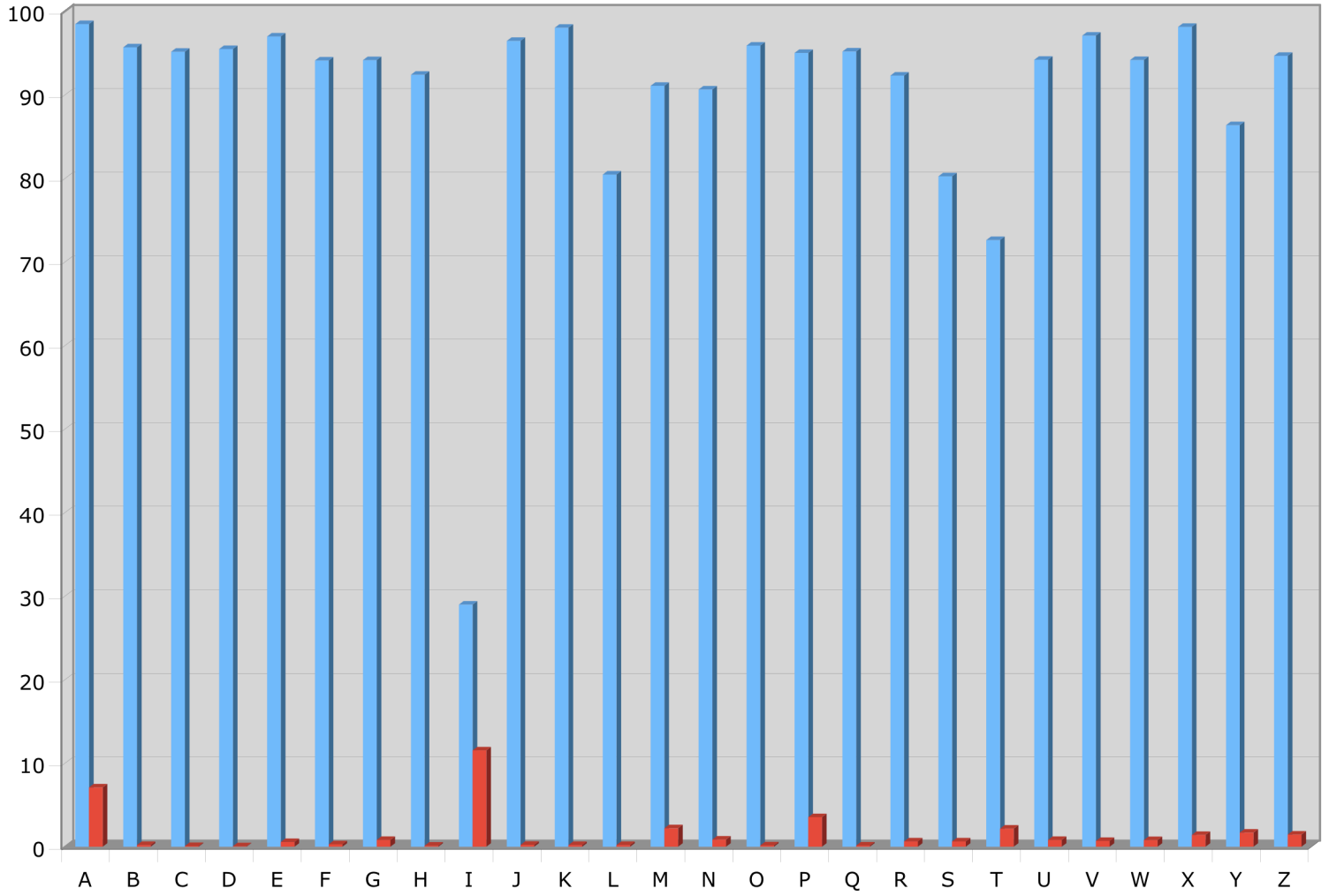
Myth 1:
All Products Are The Same

- **More specifically: "All Spam Engines are the same. You should buy based on features."**
- **Why? Lots of reasons…**

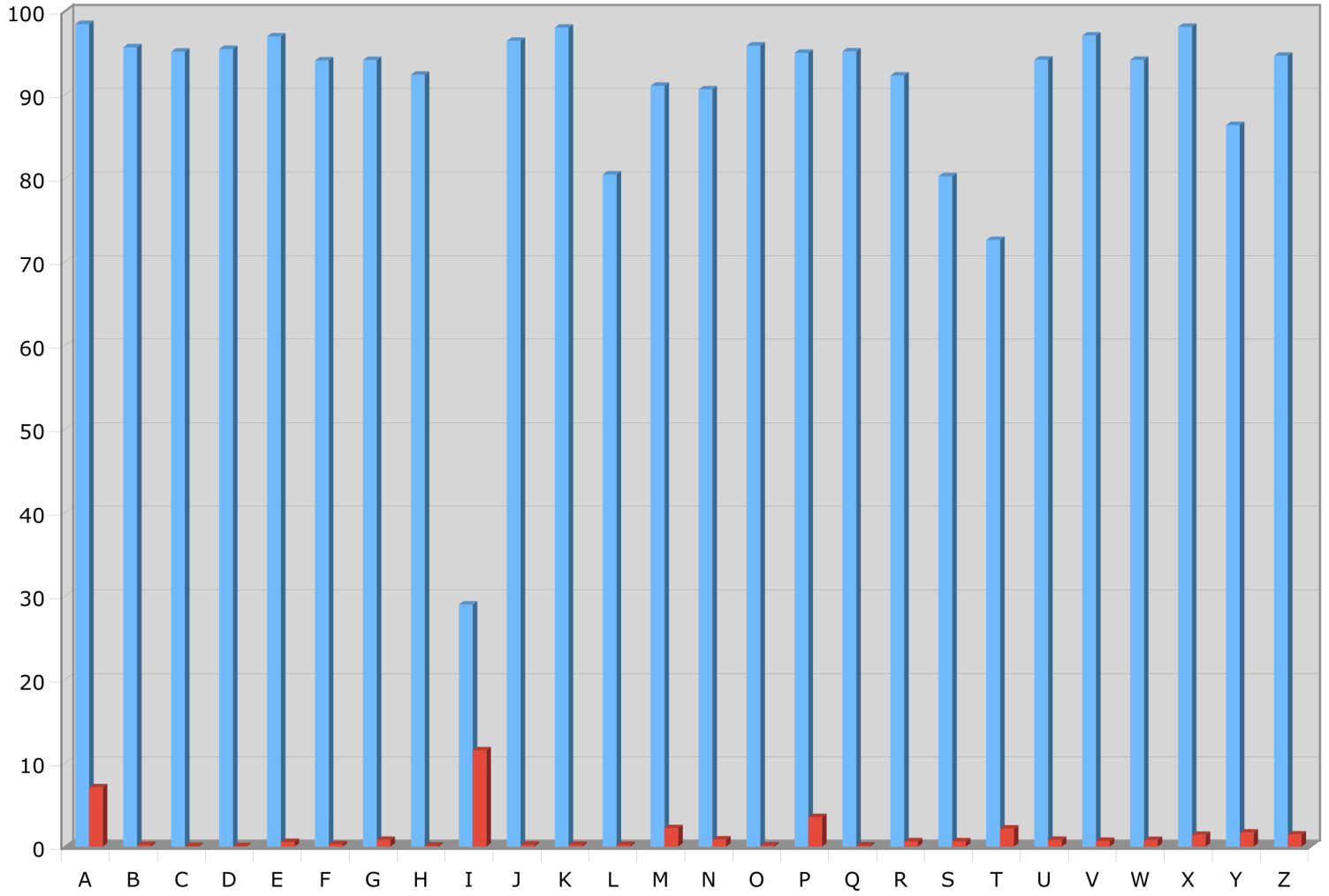
Reality: Different Spam Engines Perform Differently

Most recent results for 26 different products



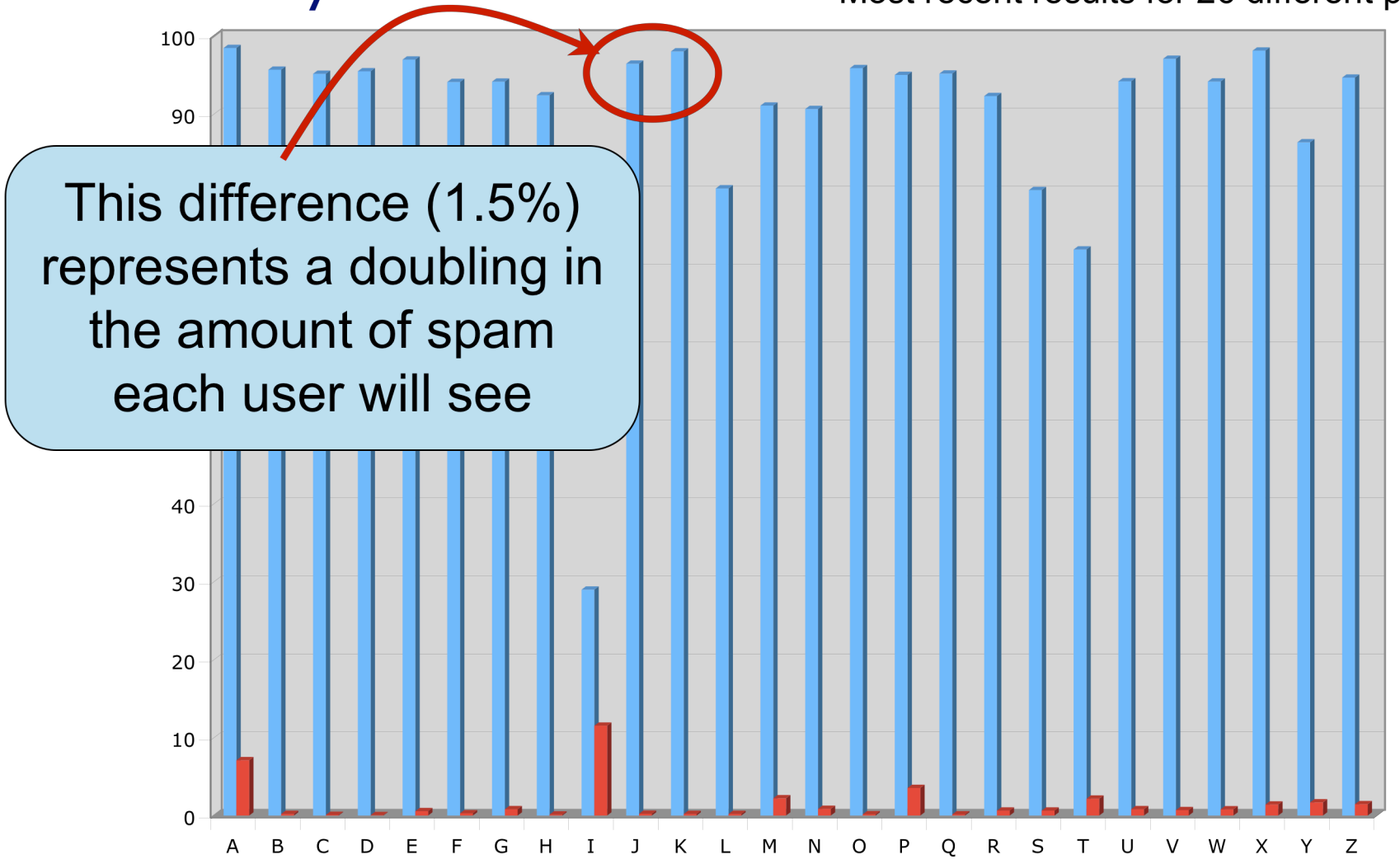
Reality: Different Spam Engines Perform Differently

Most recent results for 26 different products



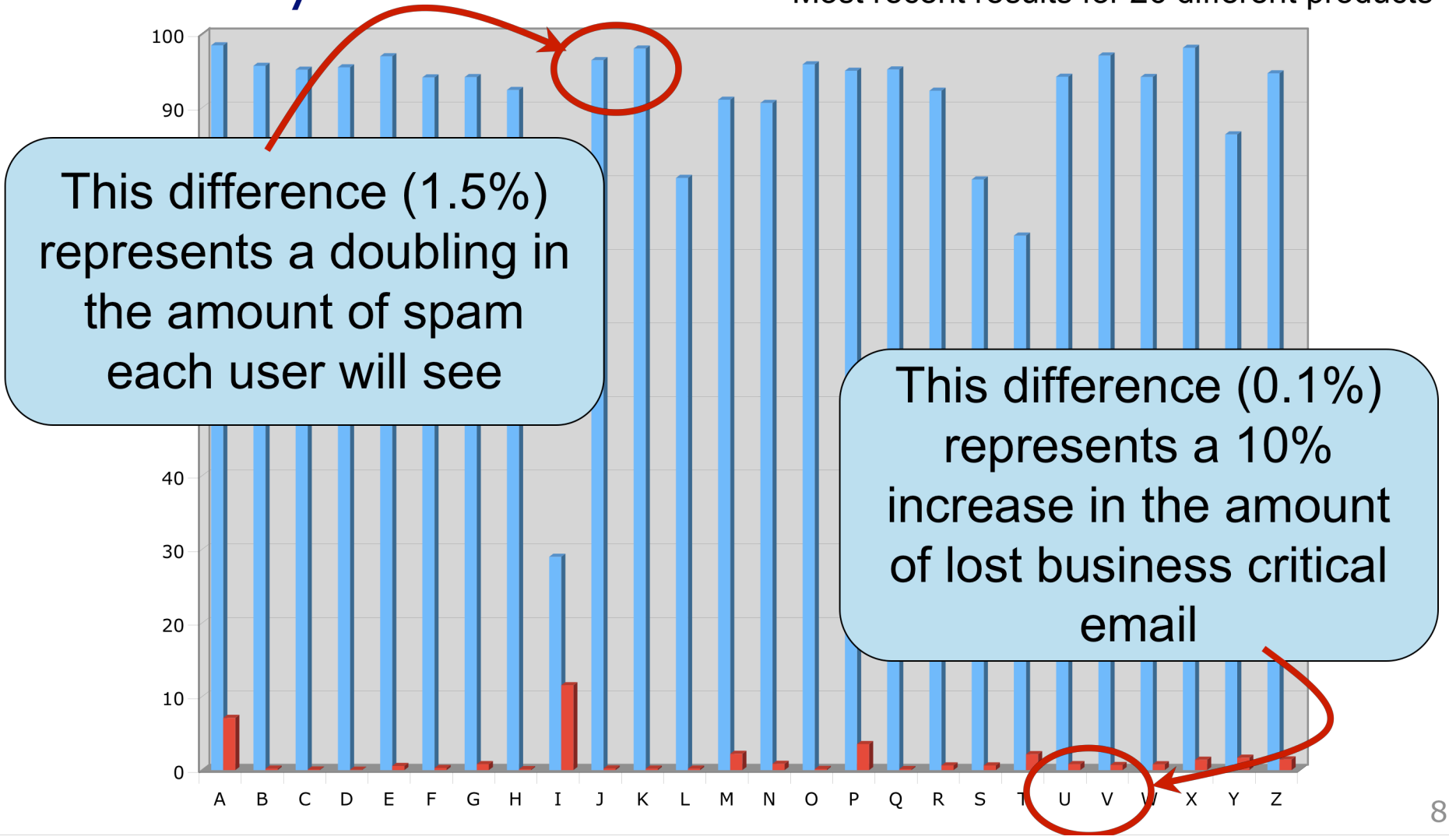
Reality: Different Spam Engines Perform Differently

Most recent results for 26 different products



Reality: Different Spam Engines Perform Differently

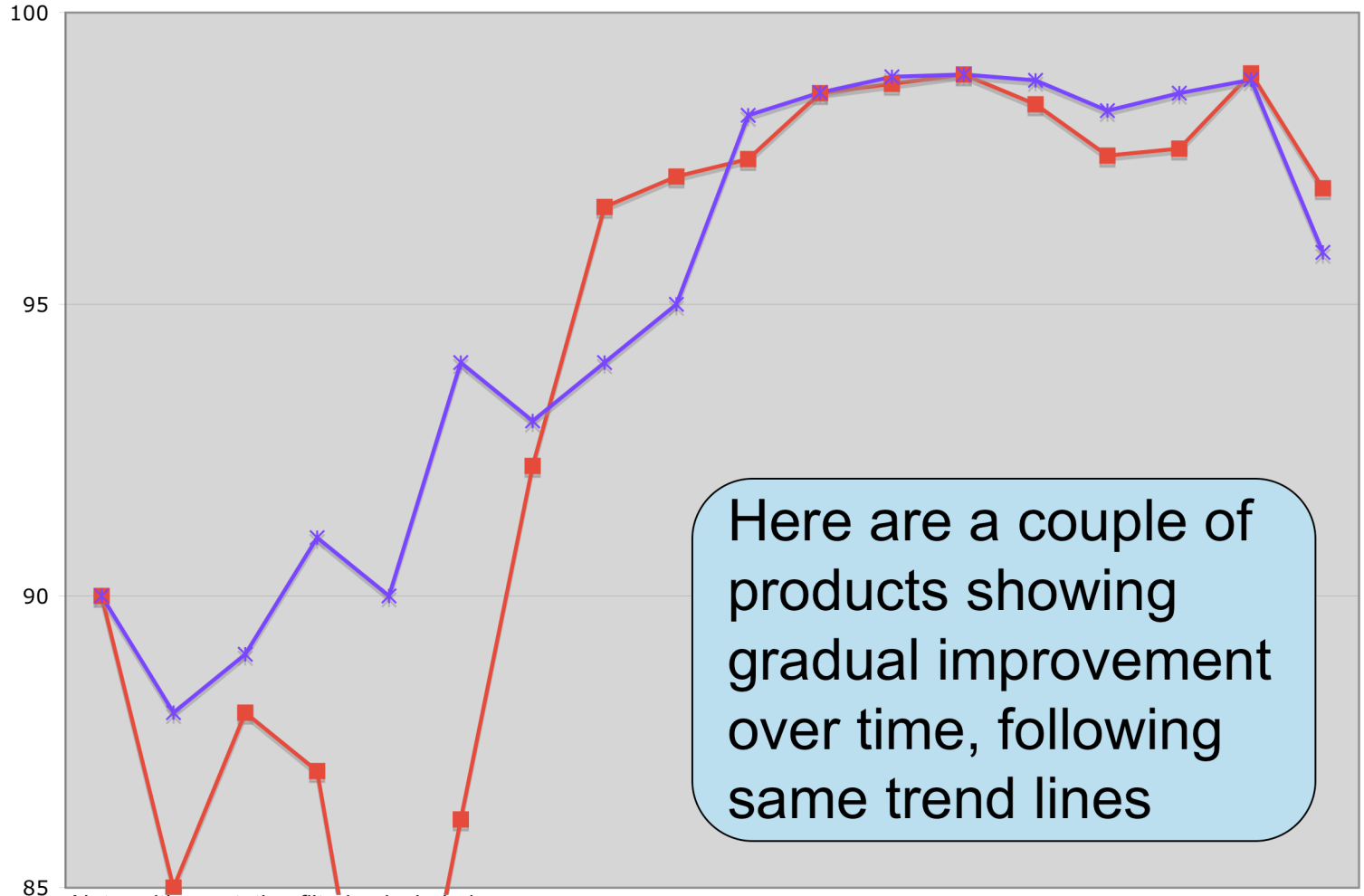
Most recent results for 26 different products



Myth 2: Great Products Stay Great

- **… And Bad Products Stay Bad**

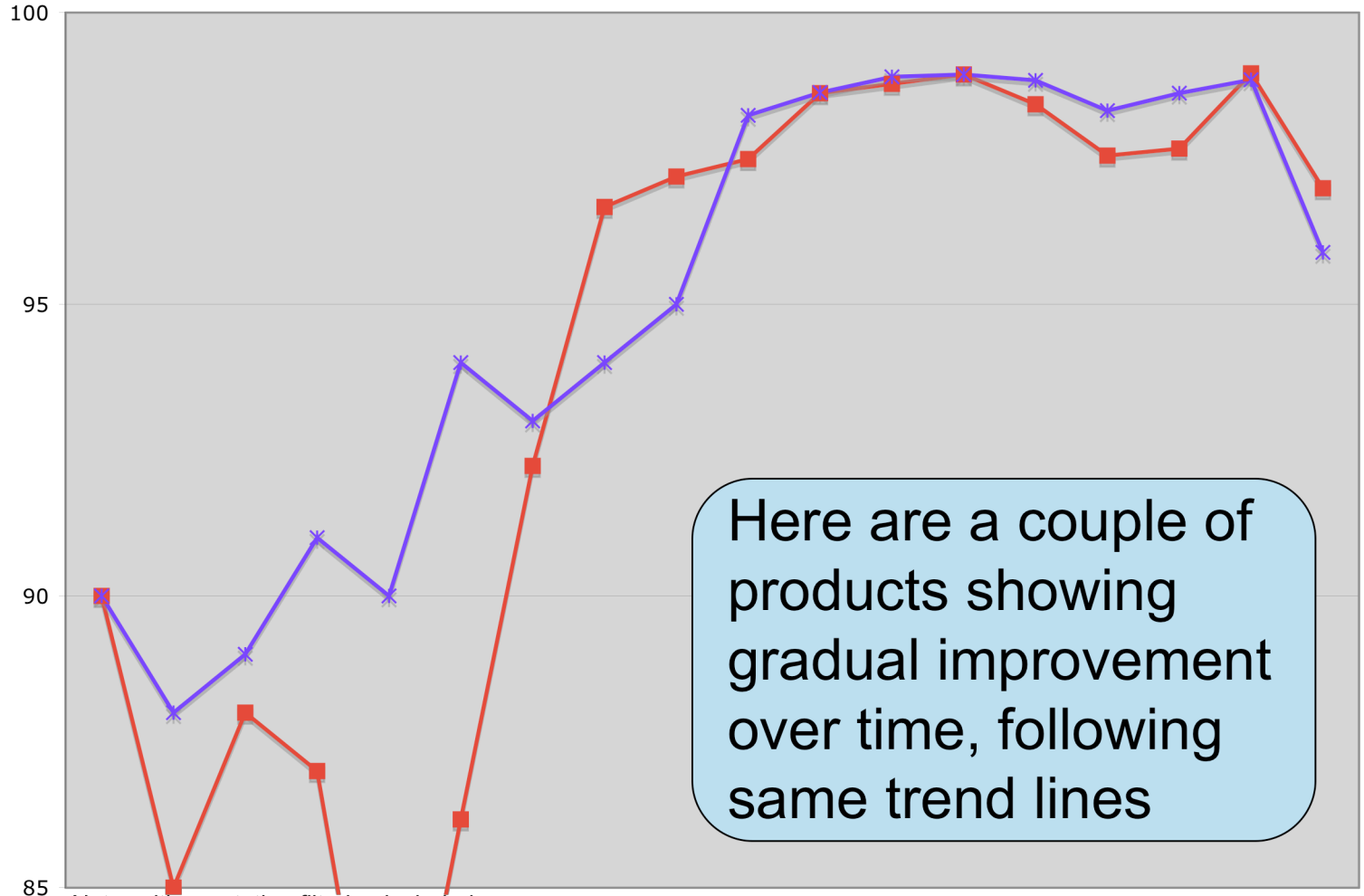
Reality:
Each Engine Has Its Own Lifecycle



Note: with reputation filtering included

Note: with reputation filtering included

Reality:
Each Engine Has Its Own Lifecycle



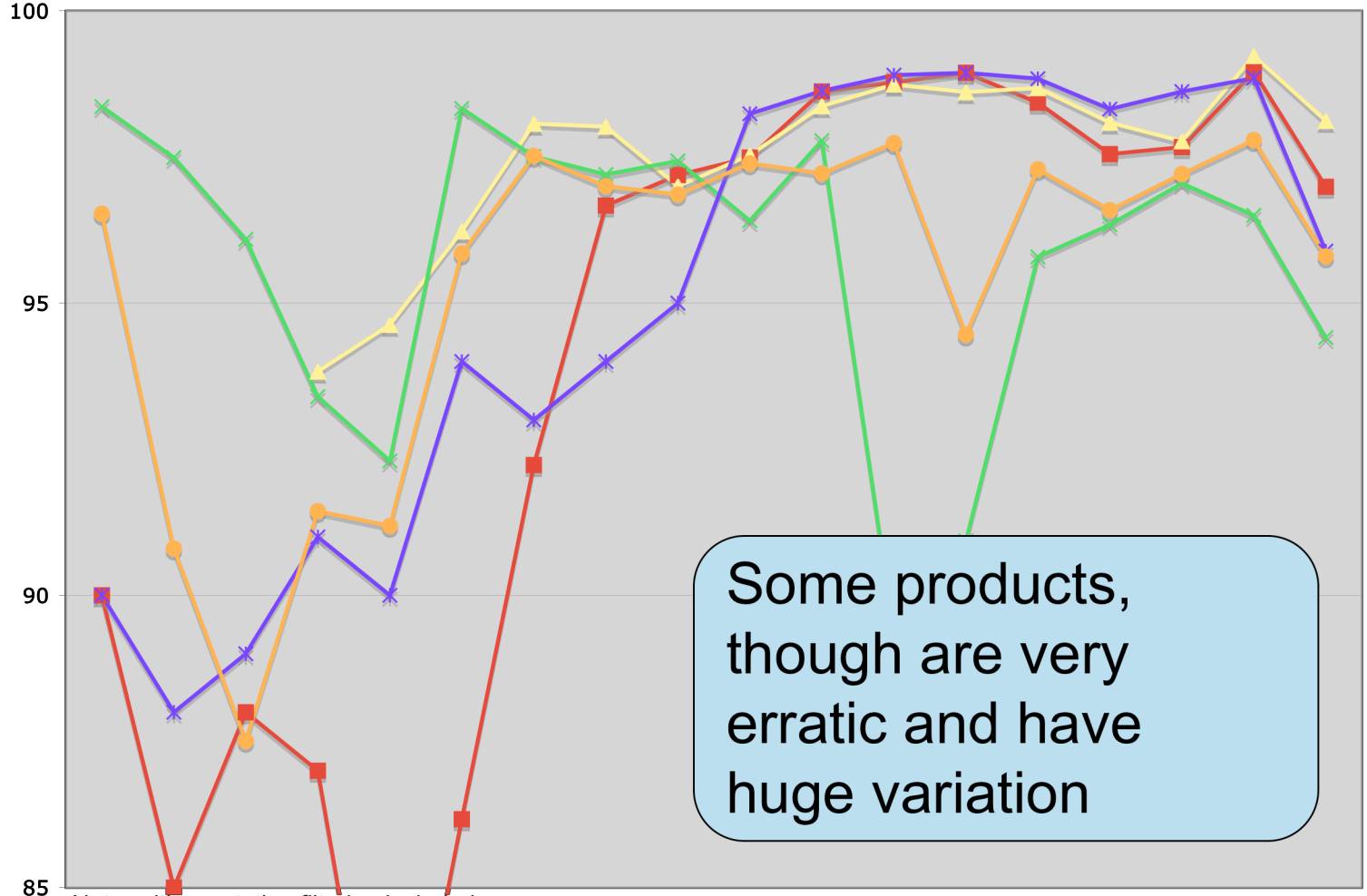
Note: with reputation filtering included

Note: with reputation filtering included

Reality:
Each Engine Has Its Own Lifecycle

Note: with reputation filtering included
Note: with reputation filtering included

Reality:
Each Engine Has Its Own Lifecycle



Note: with reputation filtering included
Note: with reputation filtering included

Reality:
Each Engine Has Its Own Lifecycle

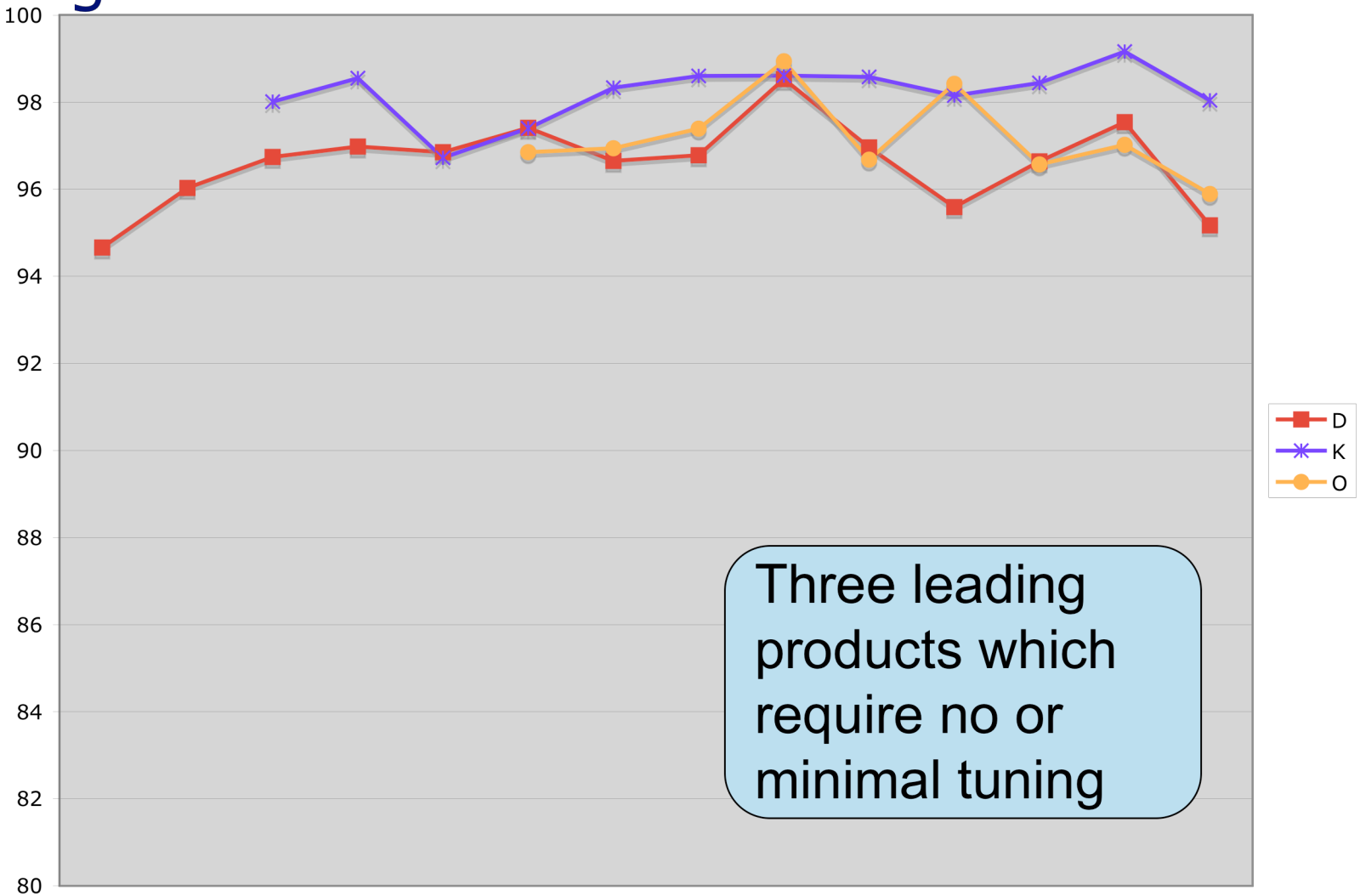
Note: with reputation filtering included
Note: with reputation filtering included

Myth 3: Training And Tuning Products Is Good

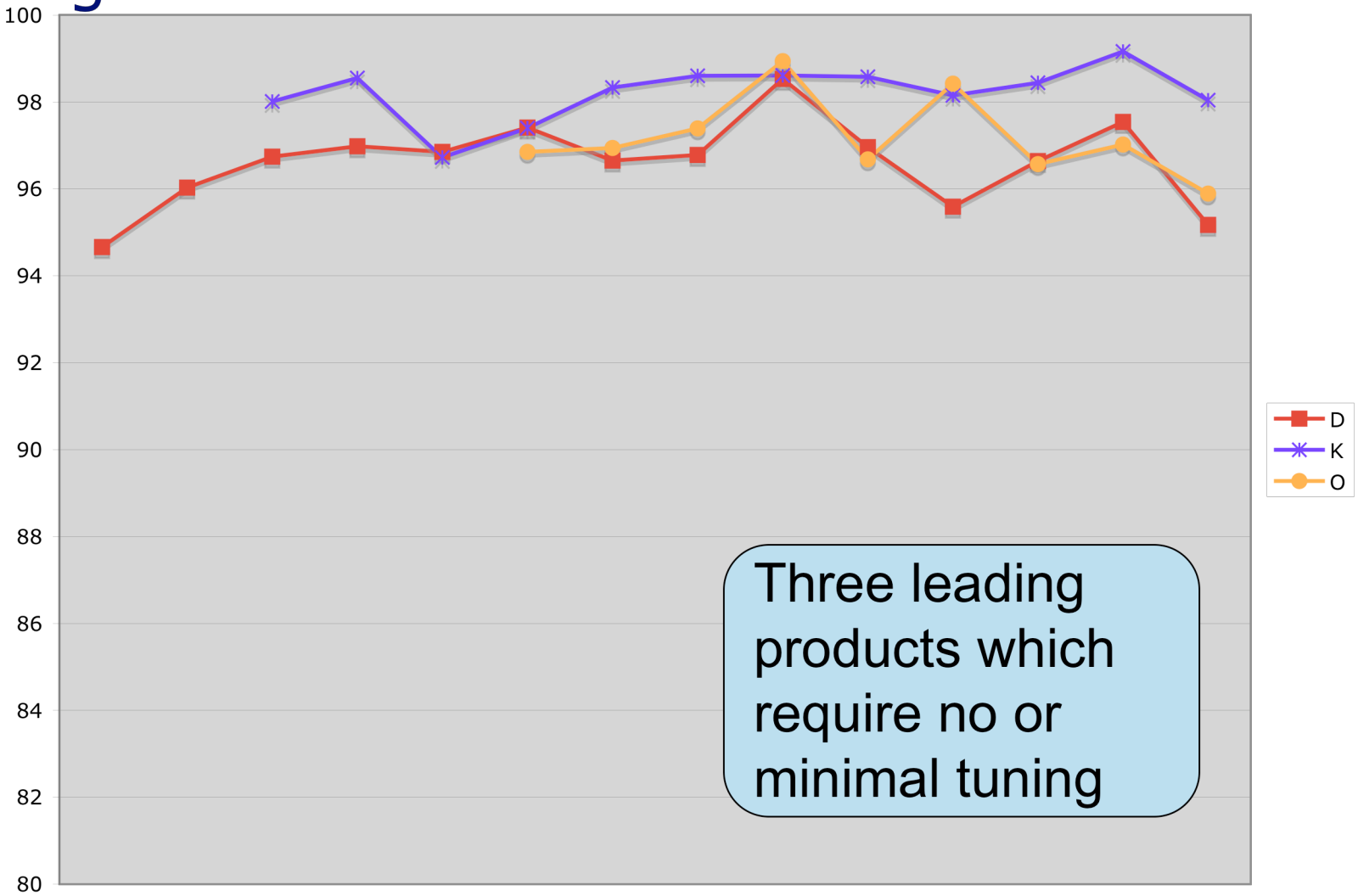
- **Where does this one come from?**
  - **Technology Guru installs Spam Assassin**
  - **Spam Assassin requires training**
  - **Technology Guru is happy**
  - **Ergo: Spam Filtering requires Training**
  - **Q.E.D.**

Tuning is a spectrum… all products allow for some tuning. We're just differentiating between levels of adjustment.

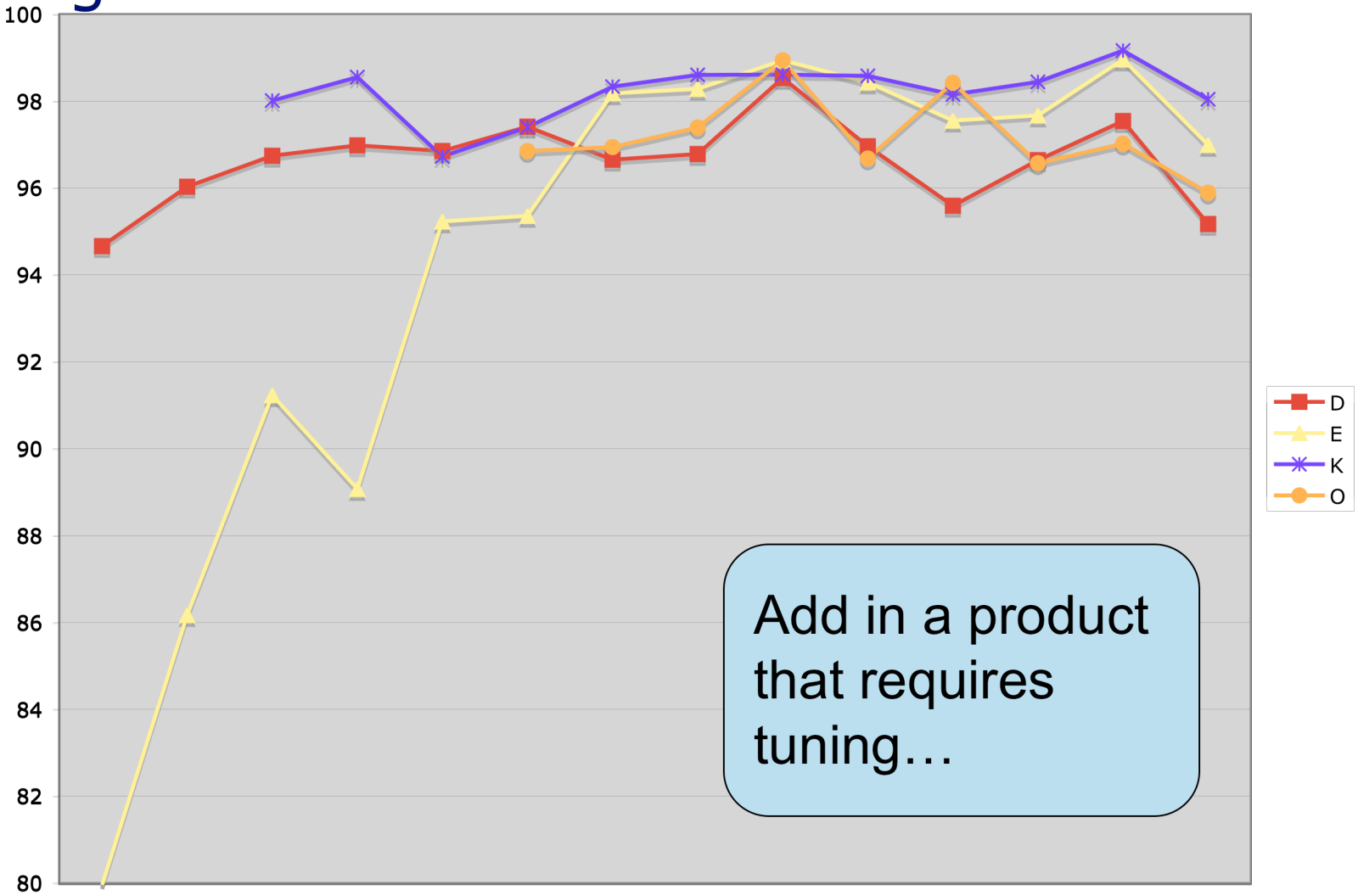
Reality:
Training Takes Time And Doesn't Work



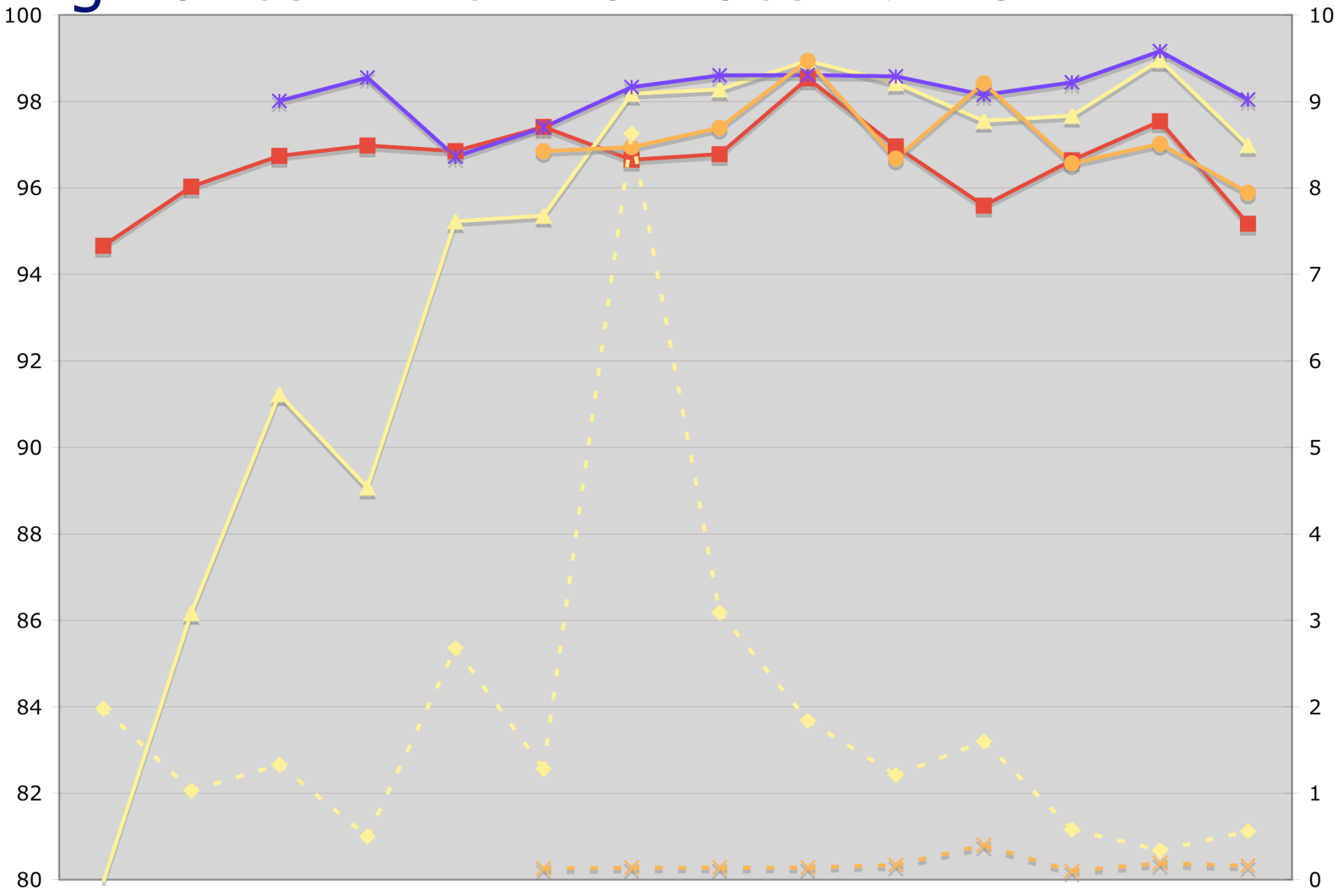
Reality:
Training Takes Time And Doesn't Work



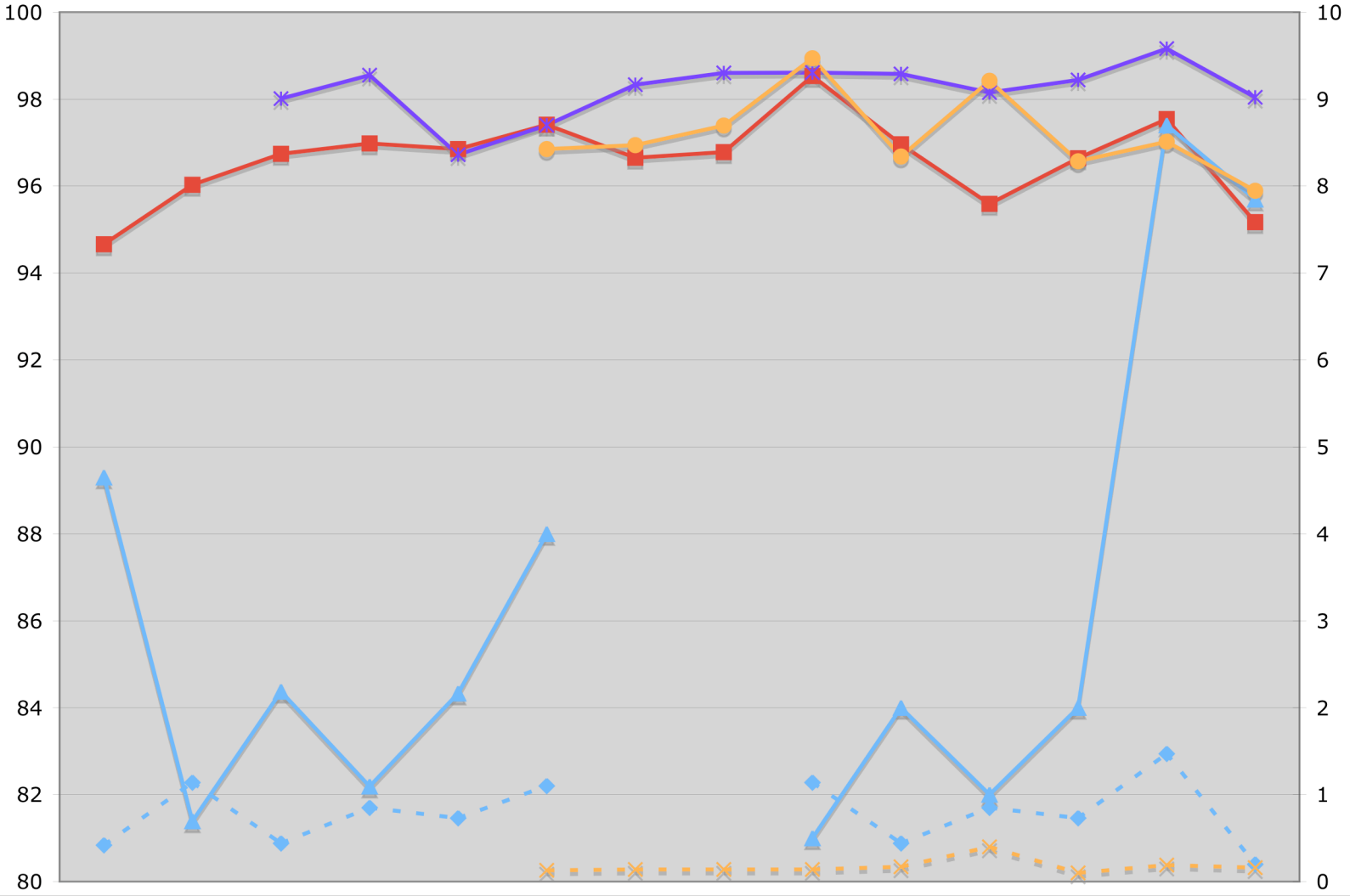
Reality:
Training Takes Time And Doesn't Work



Reality:
Training Takes Time And Doesn't Work



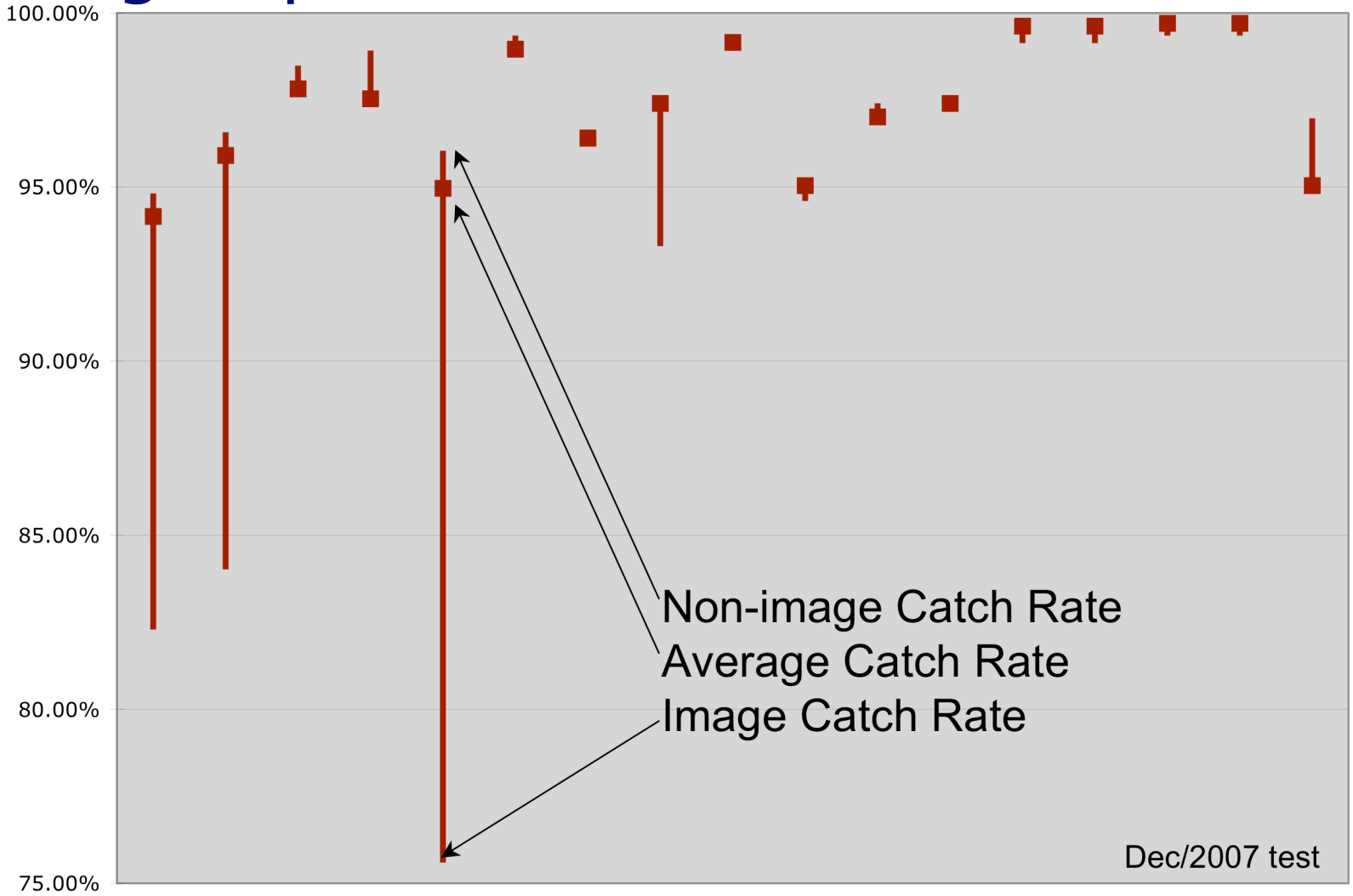
More Reality:
OK, Let's Try A Different Product



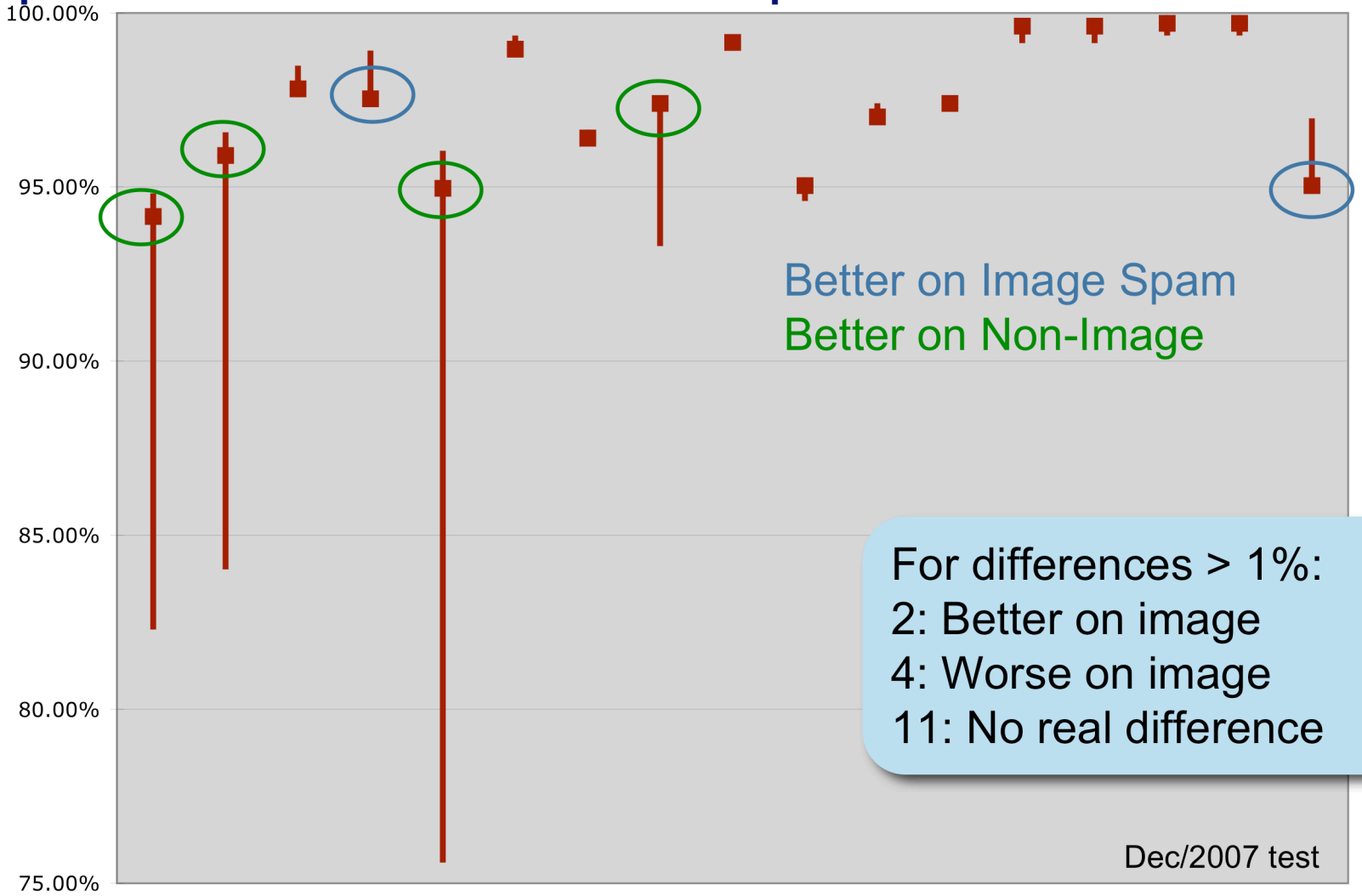
Myth 4: Image Spam Is A Problem

- **Where does this one come from?**
  - **Image spam is invented**
  - **Some products can't handle it**
  - **Strategy? Send out a press release explaining how awful life is, and maybe your customers will let you live while you fix it**

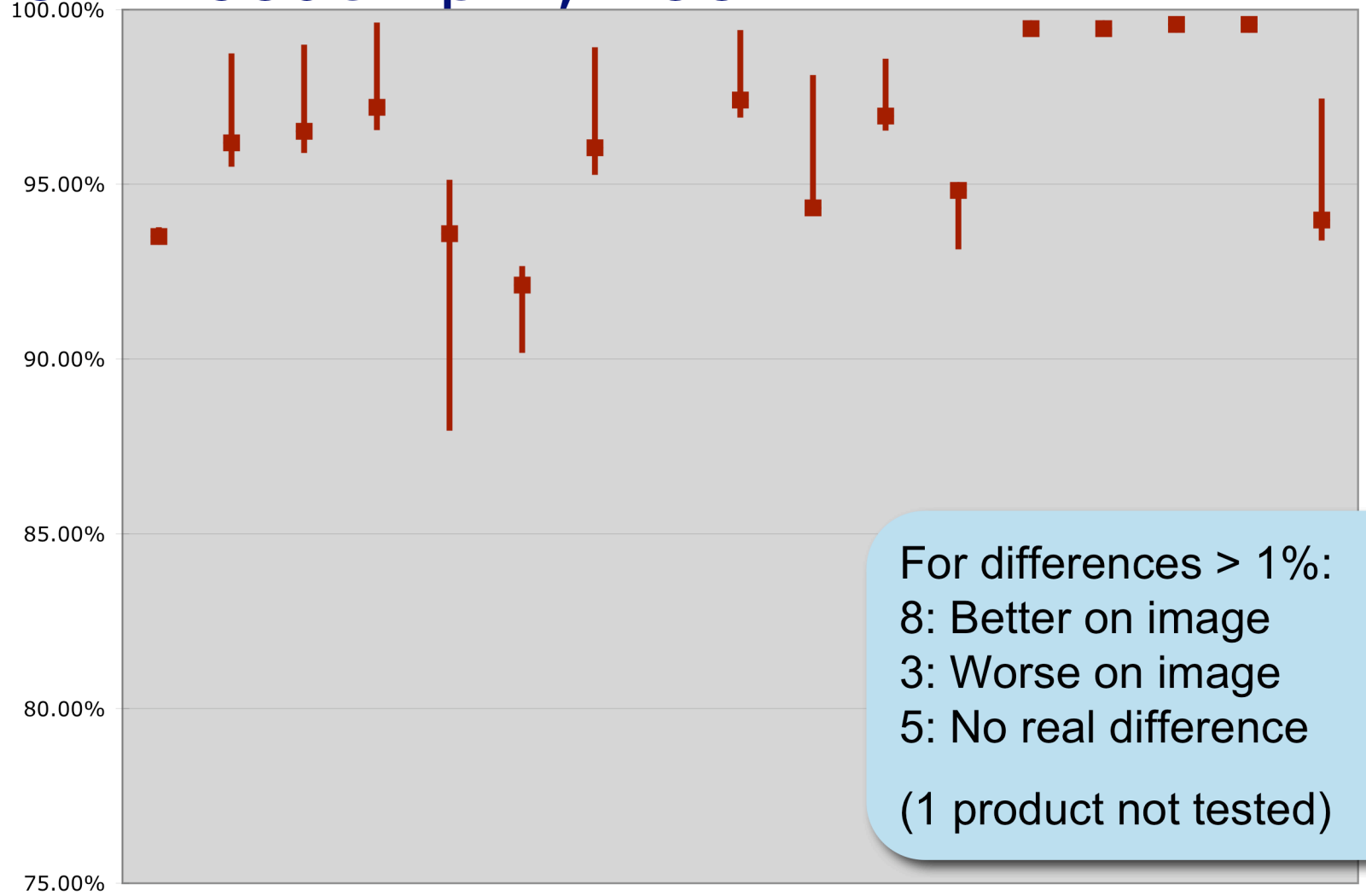
Reality:
Image Spam Is Not A Problem - For Most



In Fact, Some Products Do Better On Image Spam Than On Normal Spam



OK, You Got Me…
How About April, 2007?



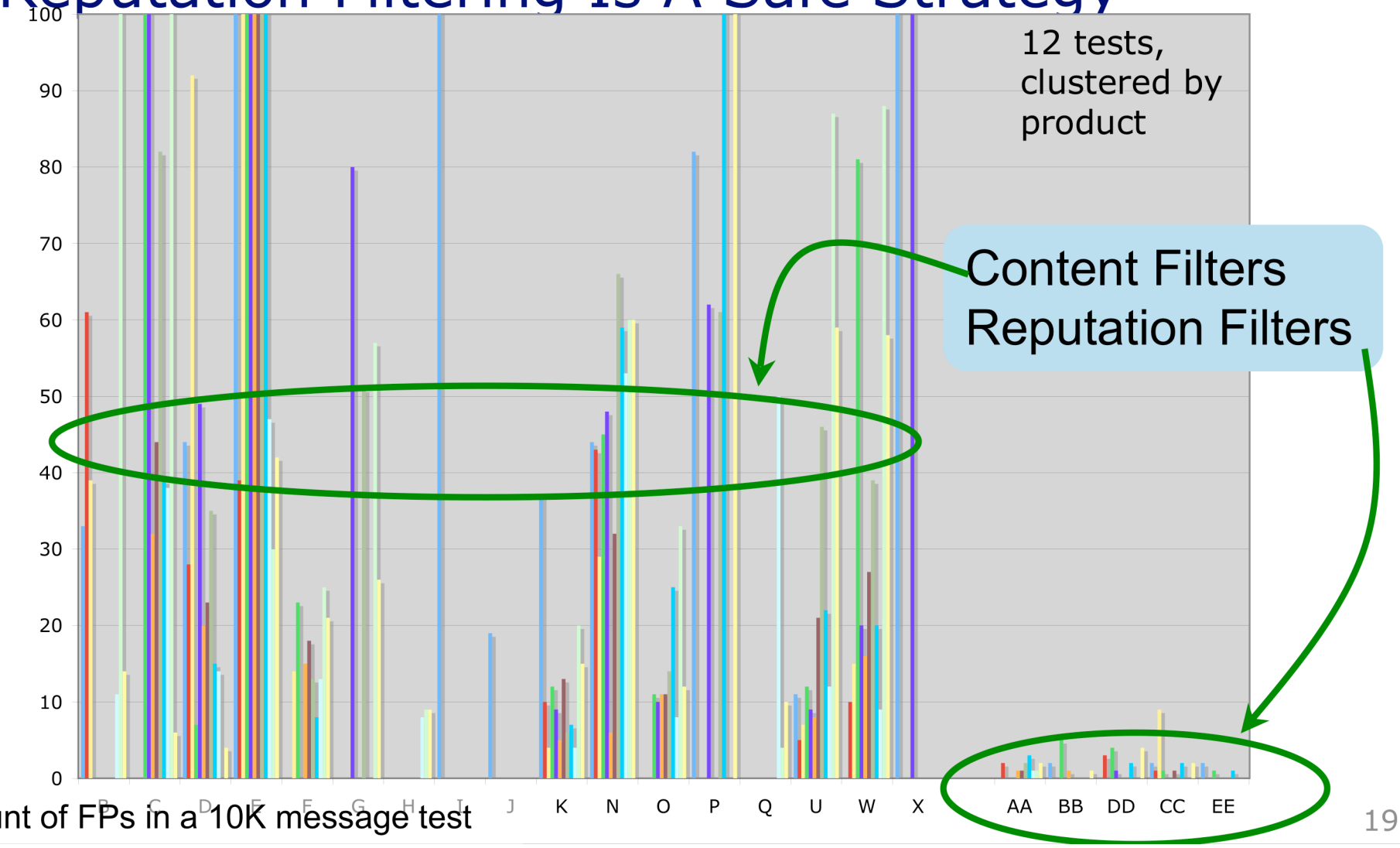
For differences > 1%:
8: Better on image
3: Worse on image
5: No real difference

(1 product not tested)

Myth 5: Reputation Filtering Is Dangerous

- **… and leads to false positives**

Reality:
Reputation Filtering Is A Safe Strategy

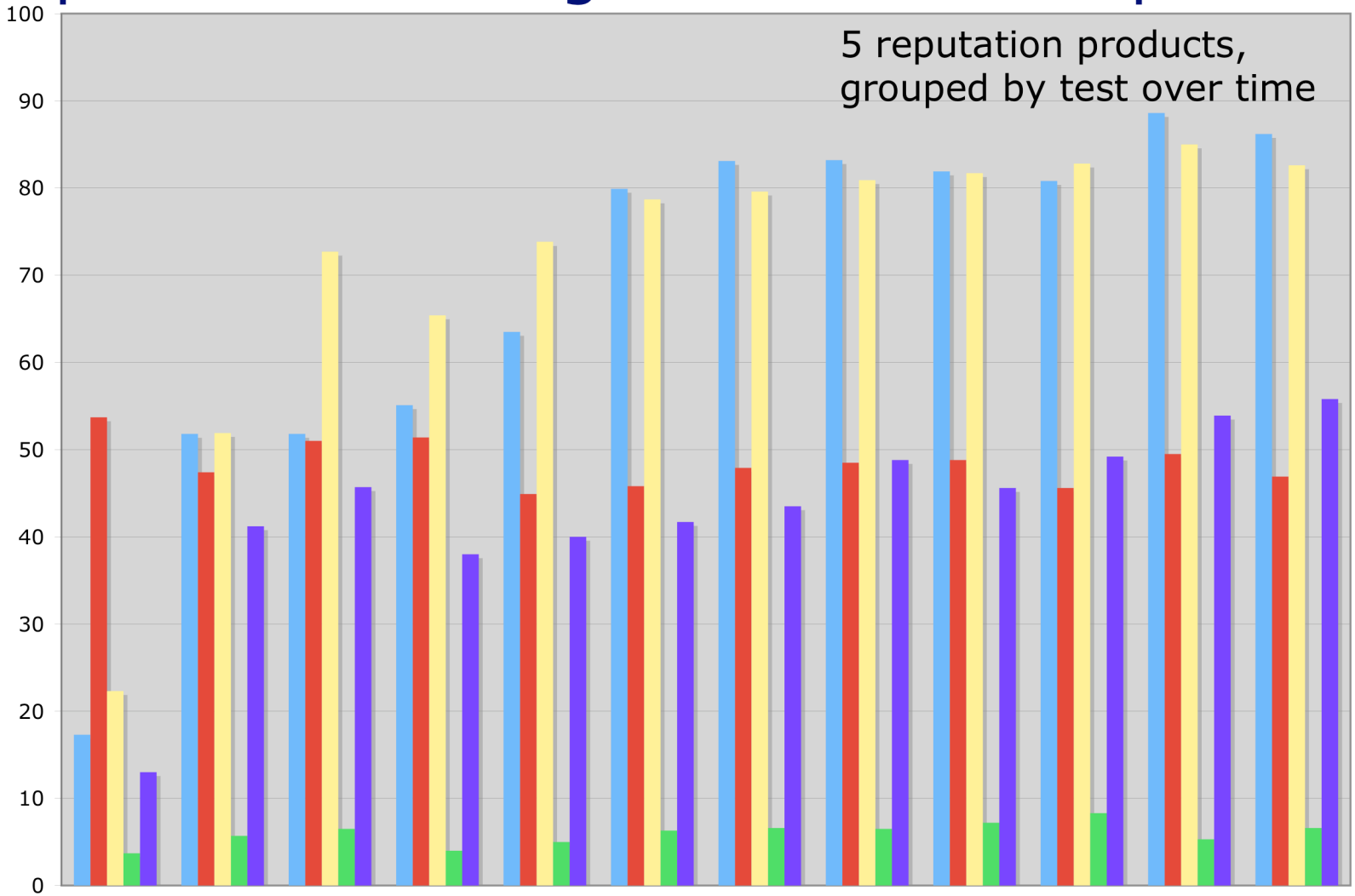


Count of FPs in a 10K message test

Myth 6: You Should Accept The Mail, And Then Apply Reputation Filtering

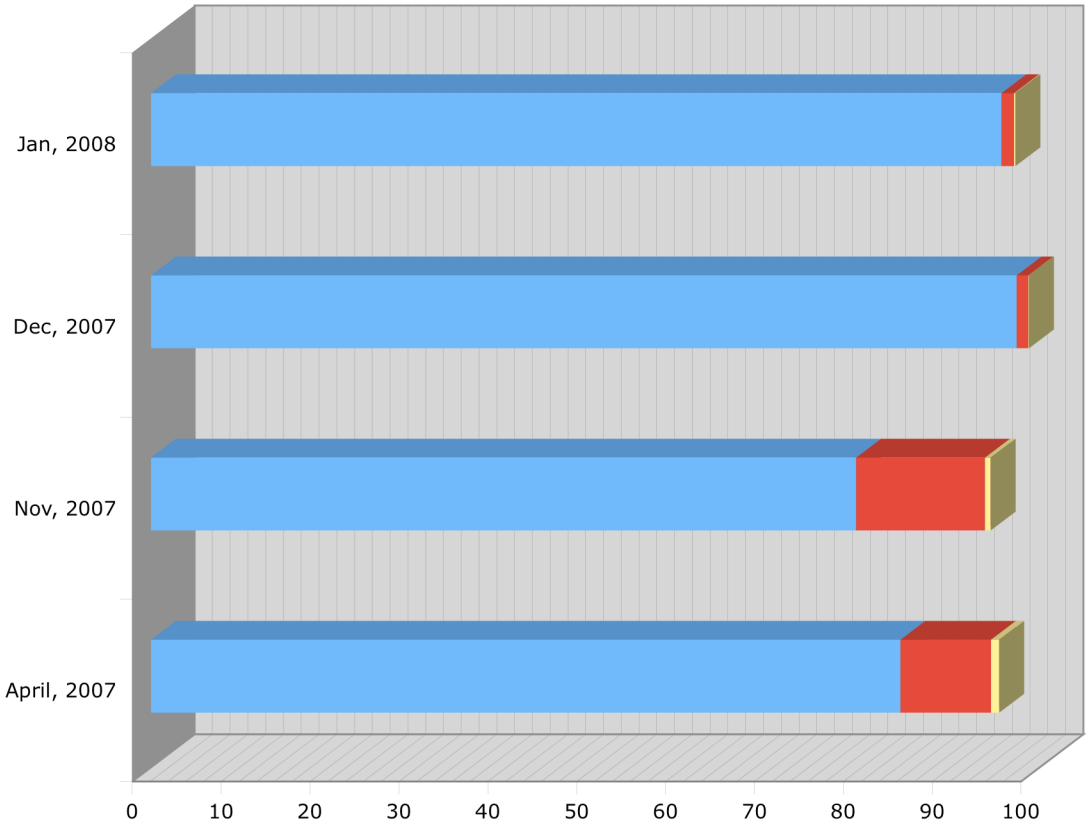
- **Where does this one come from?**
  - **Some products can't do reputation filtering because they don't have their own SMTP receiver**
  - **These products grovel through headers and simulate reputation filtering**
  - **Other products use reputation as one factor in deciding to mark mail as "Spam"**

Reality:
Reputation Filtering Blocks Lots of Spam



Interesting Side Question: Should I Stack Reputation Services?

- **After all, if one is good, aren't two better?**



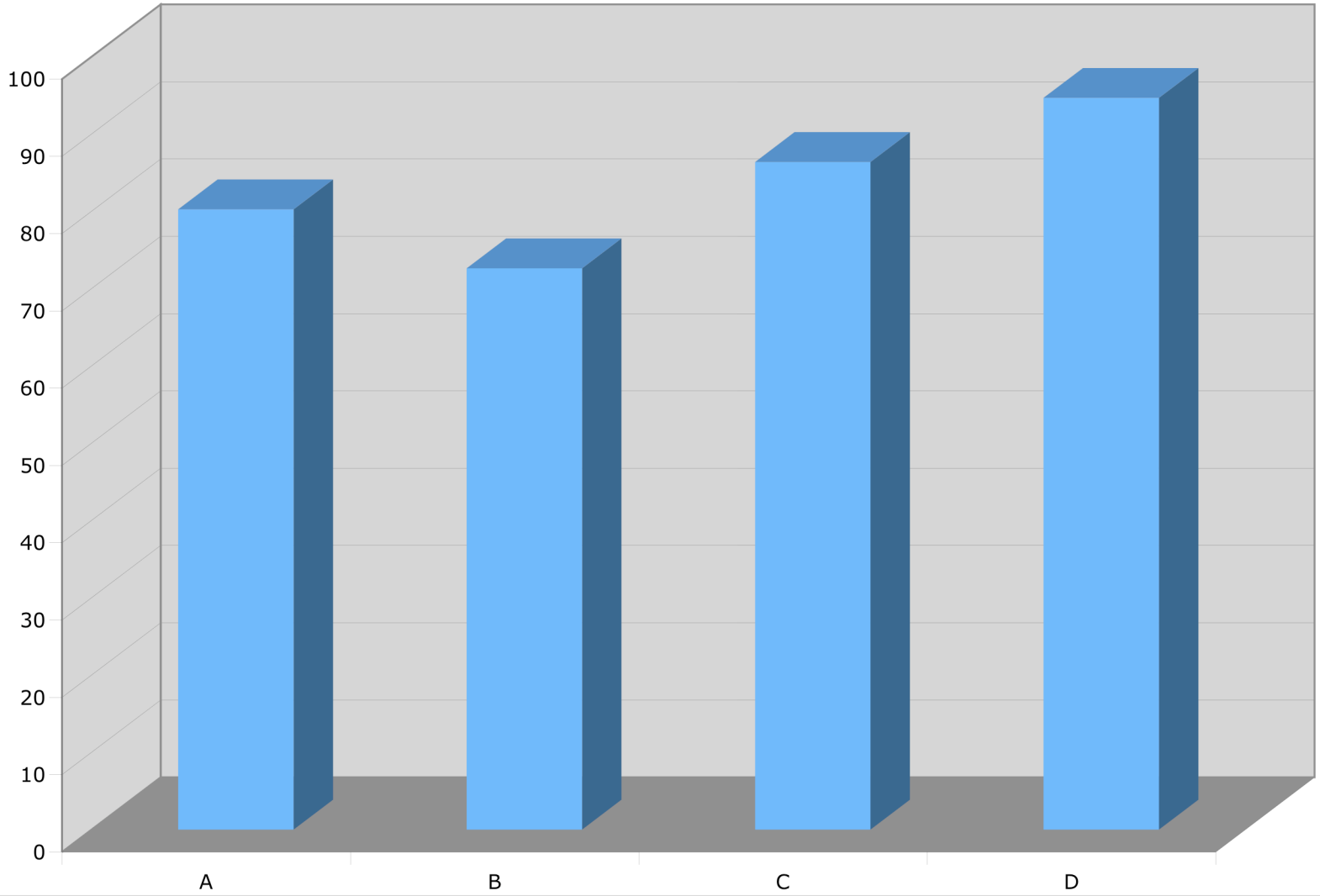
Answer: Not really worth it for traffic load and potential for false positives

But this doesn't really reflect the value of graylisting…

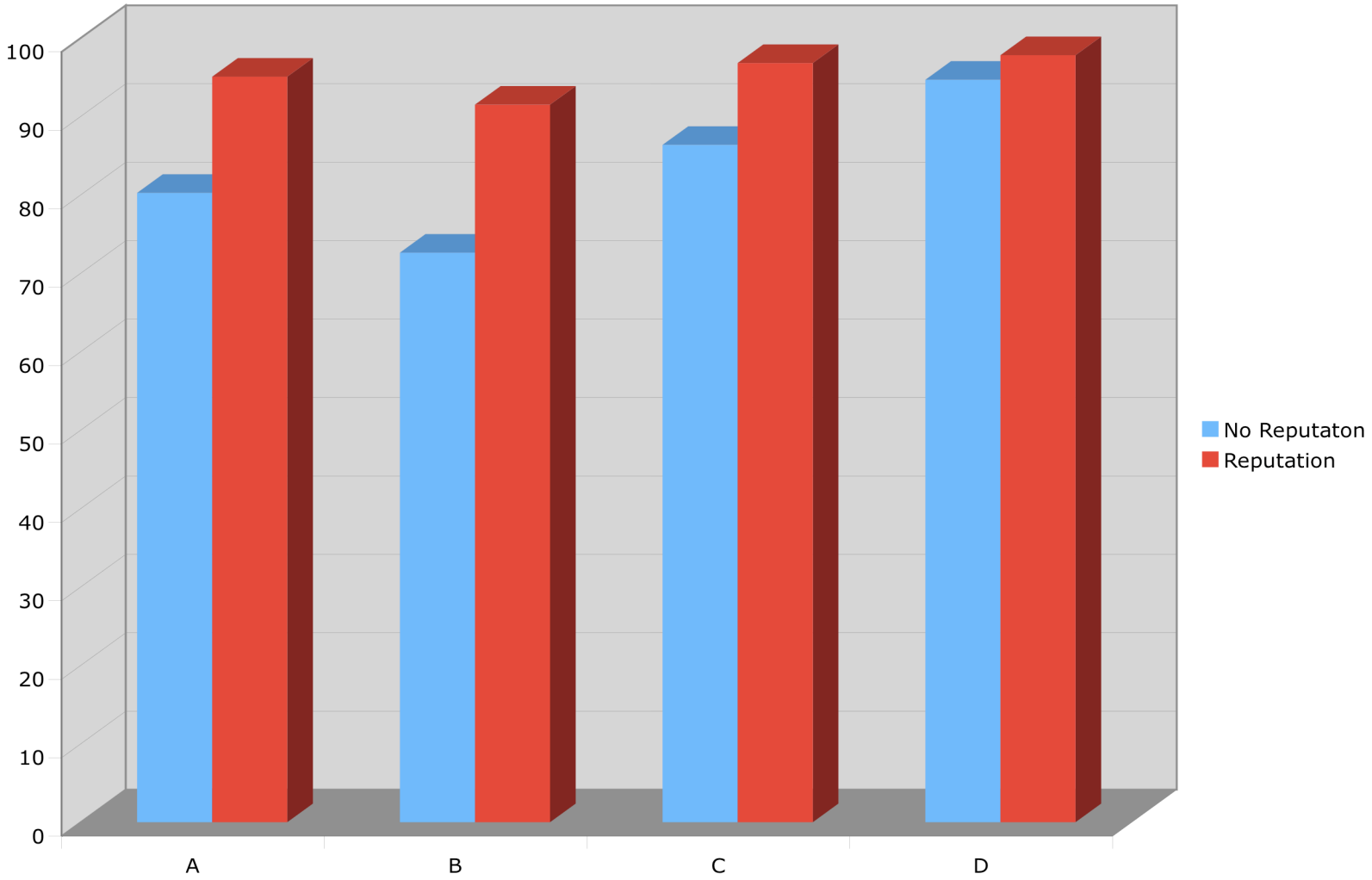
Myth 7: UTM Is A Good Way To Do Anti-Spam

- **UTM promises to make life simple by centralizing and consolidating everything into a single place**
- **Several UTM products have anti-spam built into them**
- **So this must work, right?**

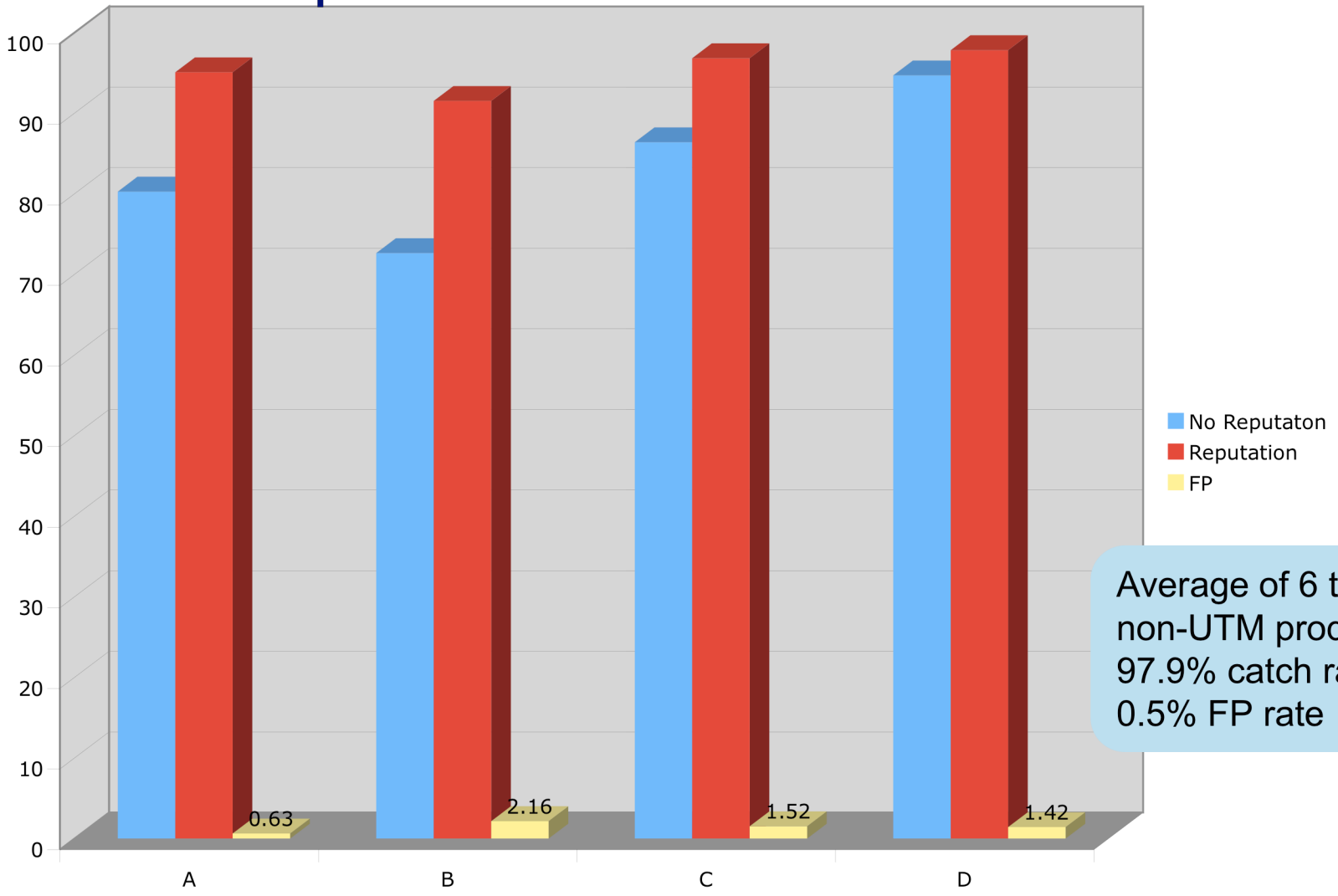
Reality:
UTM Anti-Spam As Tested Doesn't Work



Reality:
UTM Anti-Spam As Tested Doesn't Work



Reality:
UTM Anti-Spam As Tested Doesn't Work



Average of 6 tier-1 non-UTM products: 97.9% catch rate, 0.5% FP rate

Thanks!

Joel Snyder
Senior Partner
Opus One
jms@opus1.com