Lecture Outline

- Reminder: guest lecture Friday by Bill Marczak
  - Zoom link w/ password emailed out tonight
  - If you encounter difficulties, rendezvous via Piazza

- Finish botnet discussion: *Pay-per-Install* (PPI)
- Project presentations & reports
- Anonymity:
  - Brief look at Tor's evolution
    - Plus a "teachable moment"
  - Anonymizing data (packet traces)

Pay-Per-Install (PPI)

So far, Torpig has been distributed to its victims as part of Mebroot. Mebroot is a rootkit that takes control of a machine by replacing the system's Master Boot Record (MBR). This allows Mebroot

Mebroot has no malicious capability *per se*. Instead, it provides a generic platform that other modules can leverage to perform their malicious actions. In particular, Mebroot provides functionality to manage (install, uninstall, and activate) such additional modules. Immediately after the initial reboot, Mebroot contacts the *Mebroot C&C server* to obtain malicious modules (5). These modules are

Insights from the Inside: A View of Botnet Management from Infiltration

Chia Yuan Cho[§], Juan Caballero[†§], Chris Grier[§], Vern Paxson[‡§], and Dawn Song[§]
[§]UC Berkeley [†]Carnegie Mellon University [‡]ICSI

Abstract

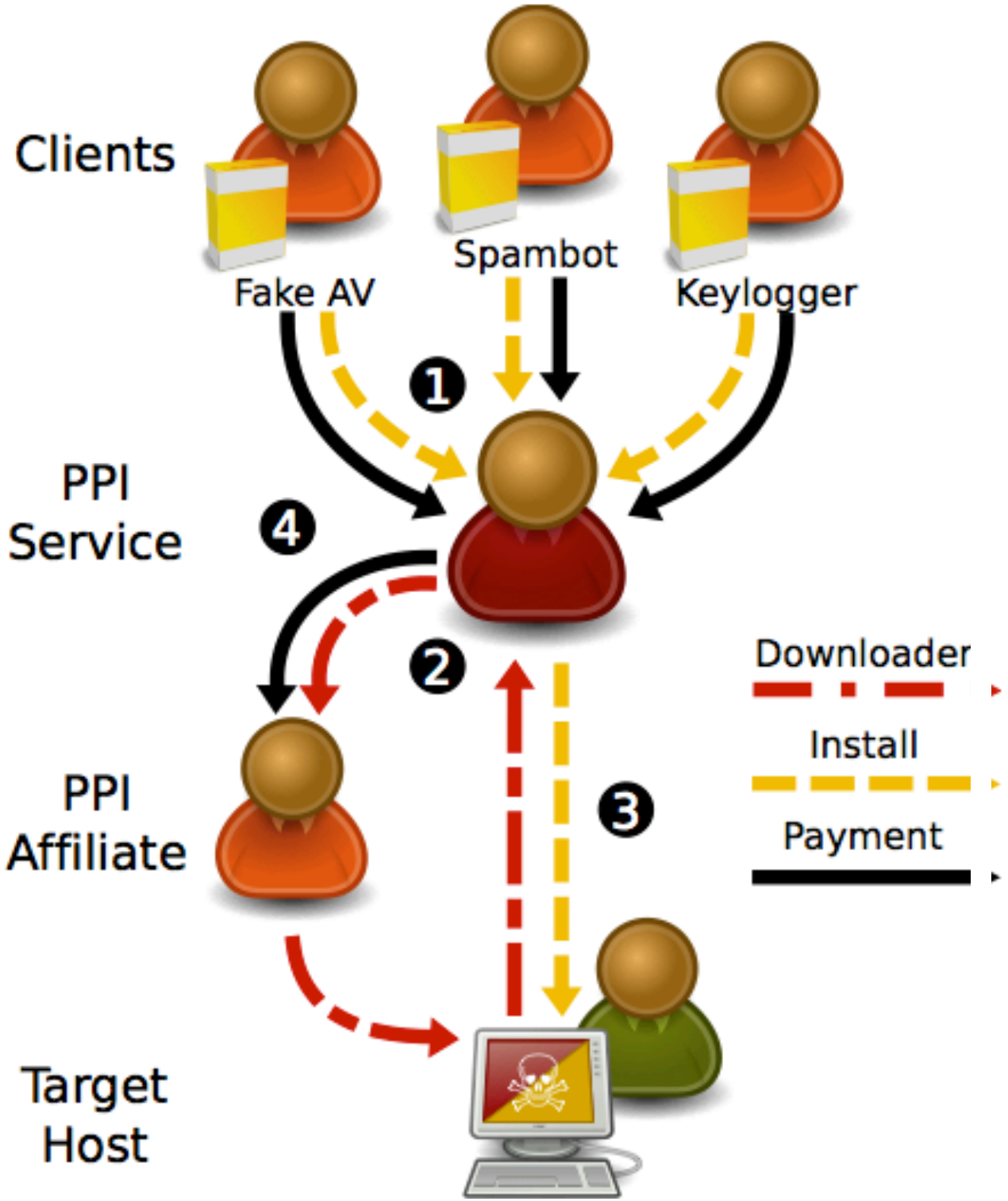
Recent work has leveraged *botnet infiltration* techniques to track the activities of bots over time, particularly with regard to spam campaigns. Building on our previous success in reverse-engineering C&C protocols, we have conducted a 4-month infiltration of the *MegaD* botnet, beginning in October 2009. Our

2009. While much of our measurement drew upon our earlier work in reverse-engineering MegaD's C&C protocol [11] and the cryptographic routines that obfuscate it [12], we also developed additional methods for gathering information about the botnet. We discovered that we could use "Google hacking" to locate additional C&C servers based on fingerprinting the web pages they sup-

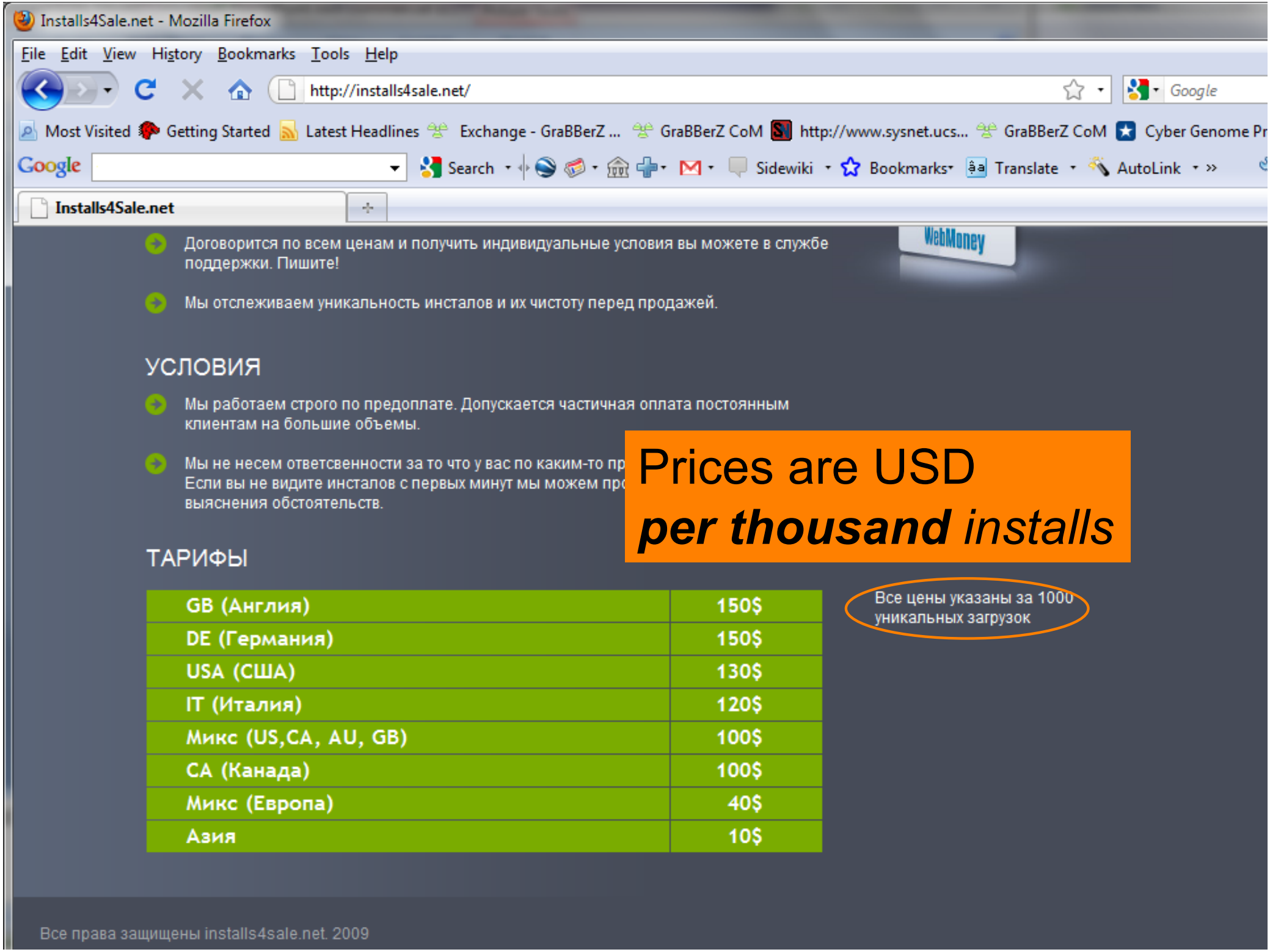
An inside view of FireEye's takedown. On Nov. 6, 2009, FireEye launched a coordinated effort to take down MegaD. The takedown was widely lauded as successful since MegaD's spam trickled to a halt. However, 16 days later its share of the world's spam exceeded its 4% pre-takedown level and by Dec. 13 it had climbed to 17% [6].


Installs4Sale.net - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://installs4sale.net/
Google
Most Visited Getting Started Latest Headlines Exchange - GraBBerZ ... GraBBerZ CoM http://www.sysnet.ucs... GraBBerZ CoM Cyber Genome Pr
Search Sidewiki Bookmarks Translate AutoLink
Installs4Sale.net
Installs4Sale.net - надежный сервис по загрузкам, достойный доверия
КОНТАКТЫ
560869831
550525933
info [at] installs4sale.net
ПРИЕМУЩЕСТВА
Быстро осуществляем отгрузку практически в любой регион. Принимаем заказы на миксы стран по вашему выбору.
Для постоянных клиентов действуют скидки и бонусы в виде дополнительного объема загрузок.
Wire
EPASS
WebMoney

The PPI Eco-system




Rus | Eng
GangstaBucks.com
Statistic
Home
Conditions
Registration
Tariffs
Contacts
An individual approach to everyone
Guaranteed weekly payouts
Round-the-clock support
Detailed statistics
User-friendly software
GangstaBucks.com - it pays on time!
We pay for all installs!
Join our ranks and by tomorrow
you could get your first payout!
GangstaBucks.com


Installs4Sale.net - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://installs4sale.net/
Most Visited Getting Started Latest Headlines Exchange - GraBBerZ ... GraBBerZ CoM http://www.sysnet.ucs... GraBBerZ CoM Cyber Genome Pr
Search Sidewiki Bookmarks Translate AutoLink
Installs4Sale.net
Договорится по всем ценам и получить индивидуальные условия вы можете в службе поддержки. Пишите!
Мы отслеживаем уникальность инсталов и их чистоту перед продажей.
WebMoney
УСЛОВИЯ
Мы работаем строго по предоплате. Допускается частичная оплата постоянным клиентам на большие объемы.
Мы не несем ответсвенности за то что у вас по каким-то пр
Если вы не видите инсталов с первых минут мы можем пр
выяснения обстоятельств.
Prices are USD
per thousand installs
ТАРИФЫ
GB (Англия) 150$
DE (Германия) 150$
USA (США) 130$
IT (Италия) 120$
Микс (US,CA, AU, GB) 100$
CA (Канада) 100$
Микс (Европа) 40$
Азия 10$
Все цены указаны за 1000 уникальных загрузок
Все права защищены installs4sale.net. 2009

Project Presentations: Logistics

- Held last two weeks of regular Semester
  - I'll finalize assignments by this weekend
- Aim for ~30 minutes of material
- Split presentation w/ partner ~50/50
- Schedule practice talk w/ me 3+ days prior
  - *Should be fully drafted and timed*
- Post short context summary to Piazza the morning before
  - *Assume the class has read it*

Project Presentations: Content

- Introduction framing apt for audience
  - A **thoughtful tour** of the problem space
    - This is the #1 value take-away for your fellow students
  - What you tackled, why it's significant
  - Assume audience has read your Piazza summary
- Sketch of related work sufficient to appreciate contribution
  - Will also address some "why didn't you try X?" questions
  - Frame how *other researchers* have undertaken evaluations in this space

Project Presentations: Content, con't

- Your strategy for pursuing your research
  - Explain technical undertaking / challenges
  - Explain evaluation methodology
- Frame the "data"
  - What does it cover
  - What does it *not* cover
  - What you know about quality/representativeness
  - If you're doing a security analysis, the "data" is your visibility into what you're analyzing
    - E.g. source code, black-box binaries, papers

Project Presentations: Content, con't

- What unexpected issues arose?
  - Emphasize **lessons learned**, not just surprises
    - Can provide valuable take-aways for other work
- Preliminary results
  - Bring out what is significant
  - Persuade us
  - Be thoughtful in data presentation (see below)
  - Illuminate limitations
- What remains
  - For your work
  - Implications / open questions for future work

Presenting Effectively: Slides

- Think creatively
- Make judicious use of color
- Avoid serif fonts
- Avoid overly busy slides
- Avoid “wall of bullets” on slide after slide
- Use animations to engage your audience
  - Keep them from peeking ahead, deciding they got it, and tuning out
  - Focus their attention by emphasizing current discussion point / downplaying non-points

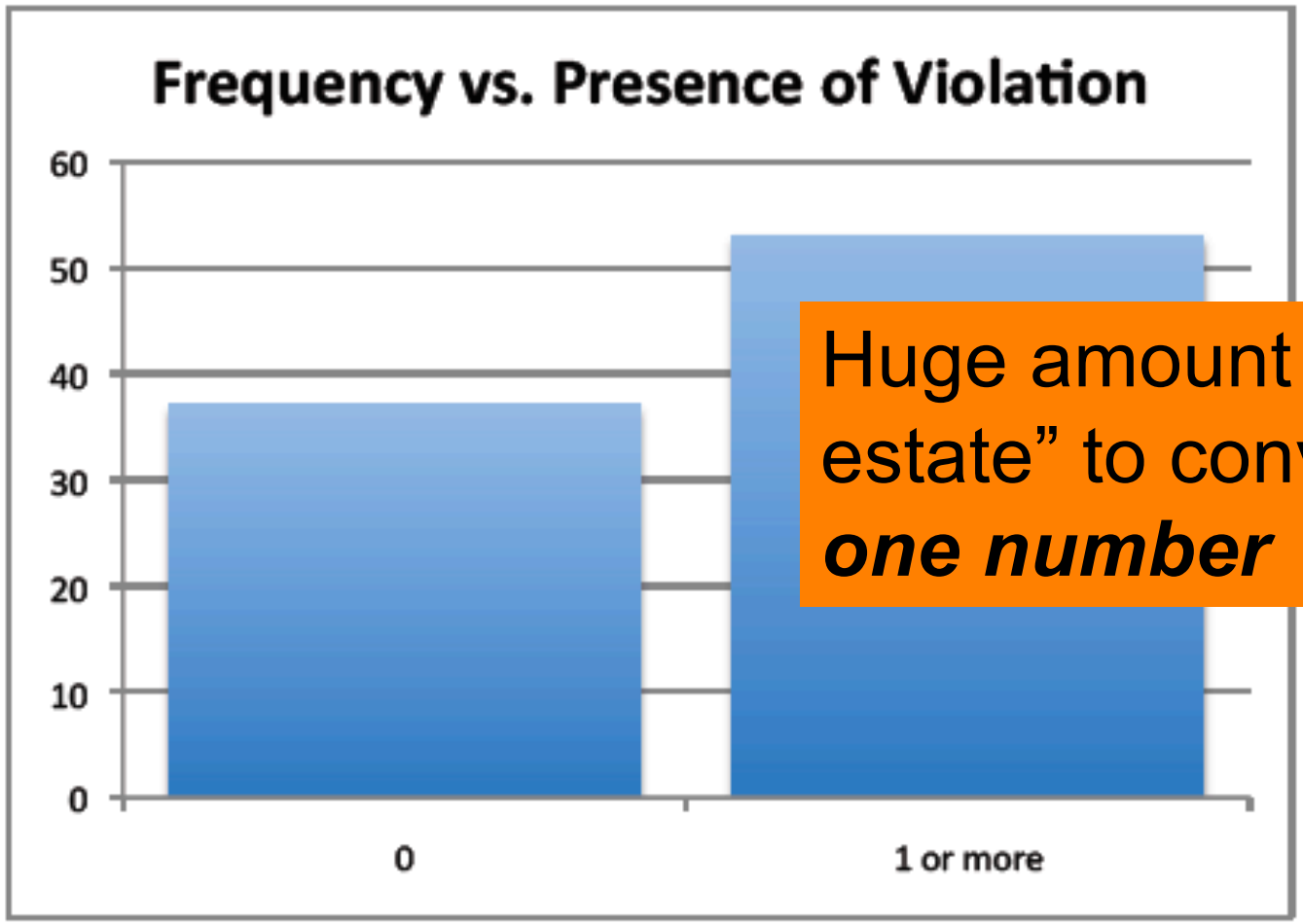
Presenting Effectively: Voice

- Do *not* read your slides
  - ProTip: short phrases force fill-ins
- Do *not* read your speaker notes
  - ProTip: try *not having any (you won't have any!)*
- Find & deliver genuine energy/enthusiasm
- Vary your tone
  - Glitches are an opportunity, not a problem: respond in the moment
- Find a conversational pace
- (Don't worry about audience eye contact!)

Project Reports

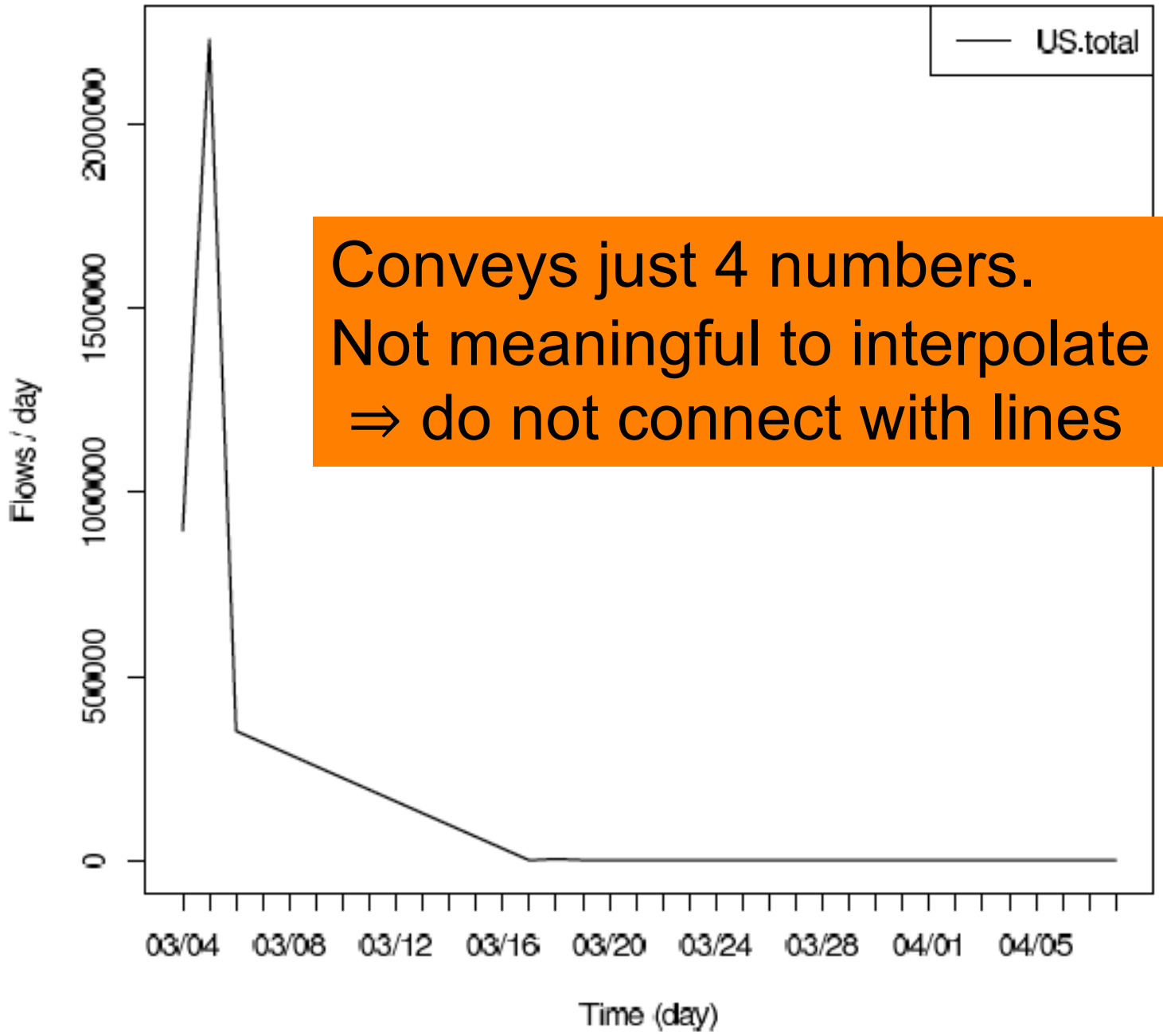
- Treat the class Projects web page as a **CFP**
  - CFP = Call For Papers
  - Formatting, deadline requirements are serious
  - Read and deliver on the Writing Pointers
    - https://www.icir.org/vern/cs261n/writing.html

- "Be thoughtful in data presentation (see below)"



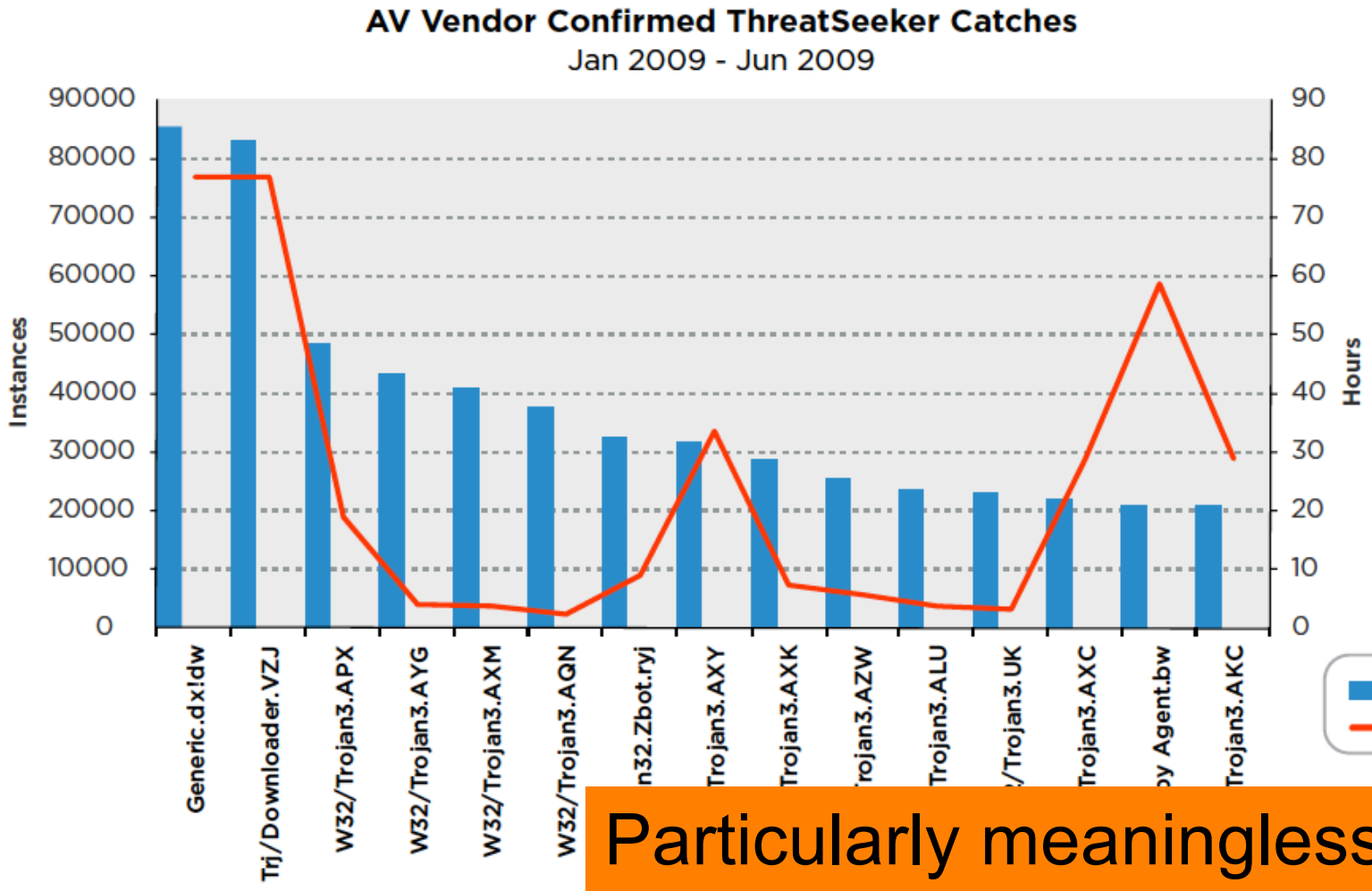
(a) Presence of violations

Huge amount of “real estate” to convey ***just one number***

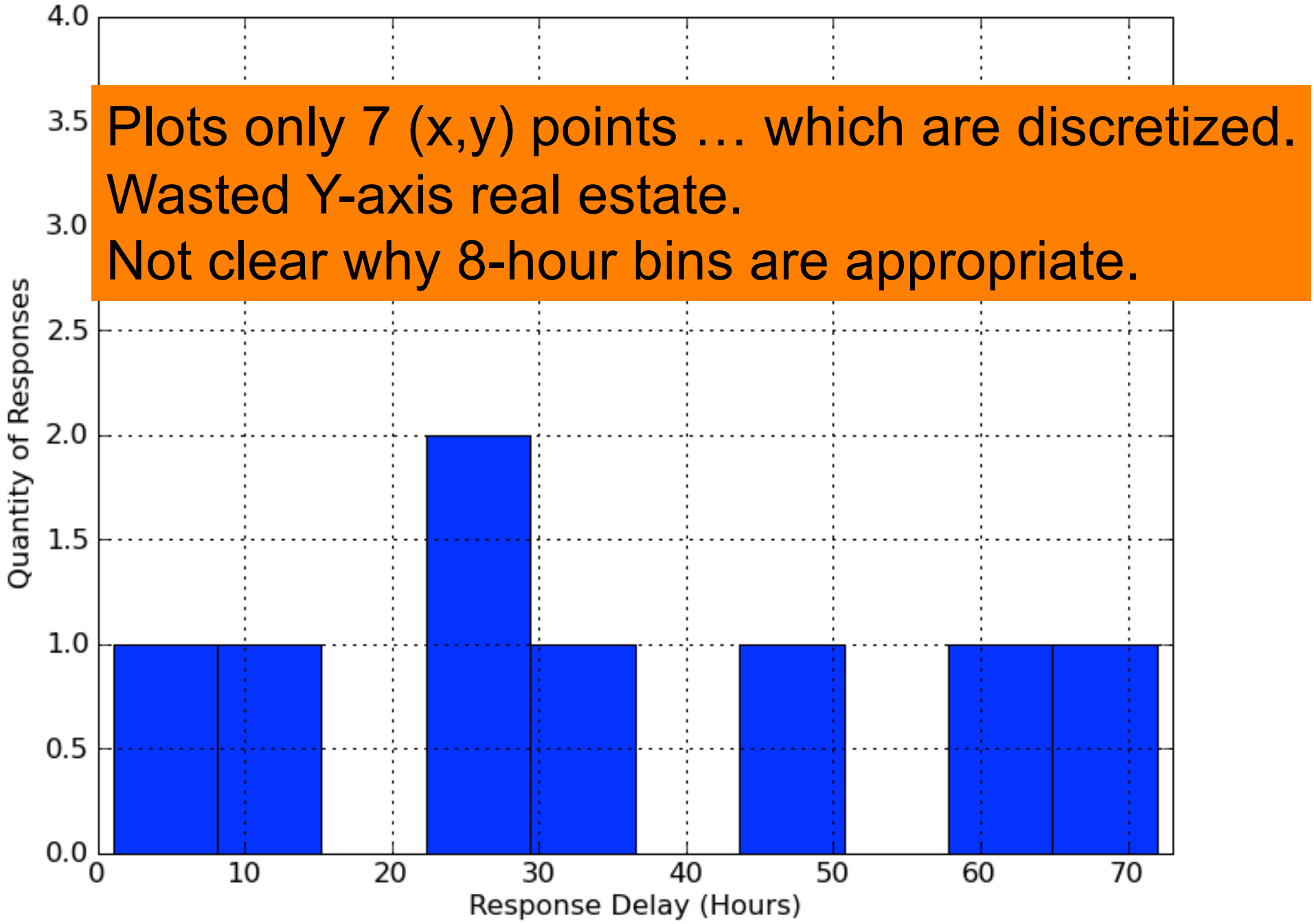


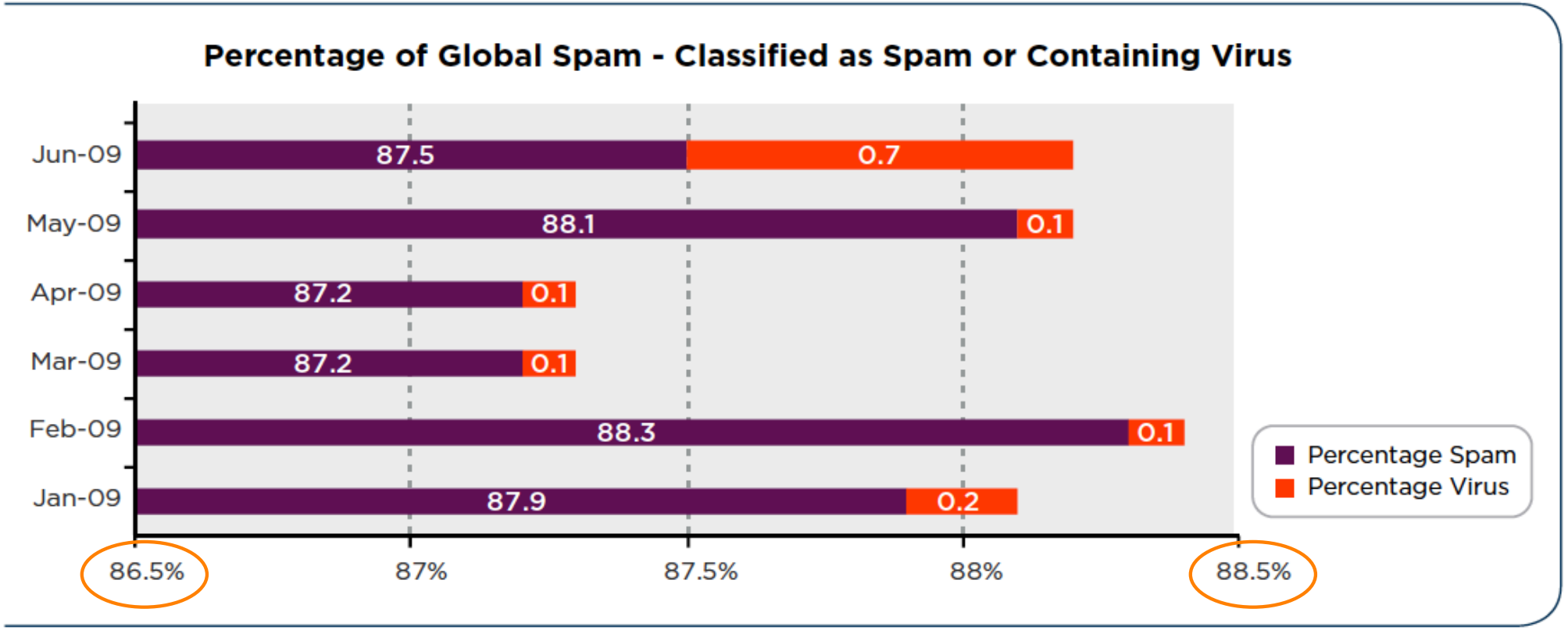
Conveys just 4 numbers.
Not meaningful to interpolate points
⇒ do not connect with lines

(b) Total traffic for `62.34.164.84`.



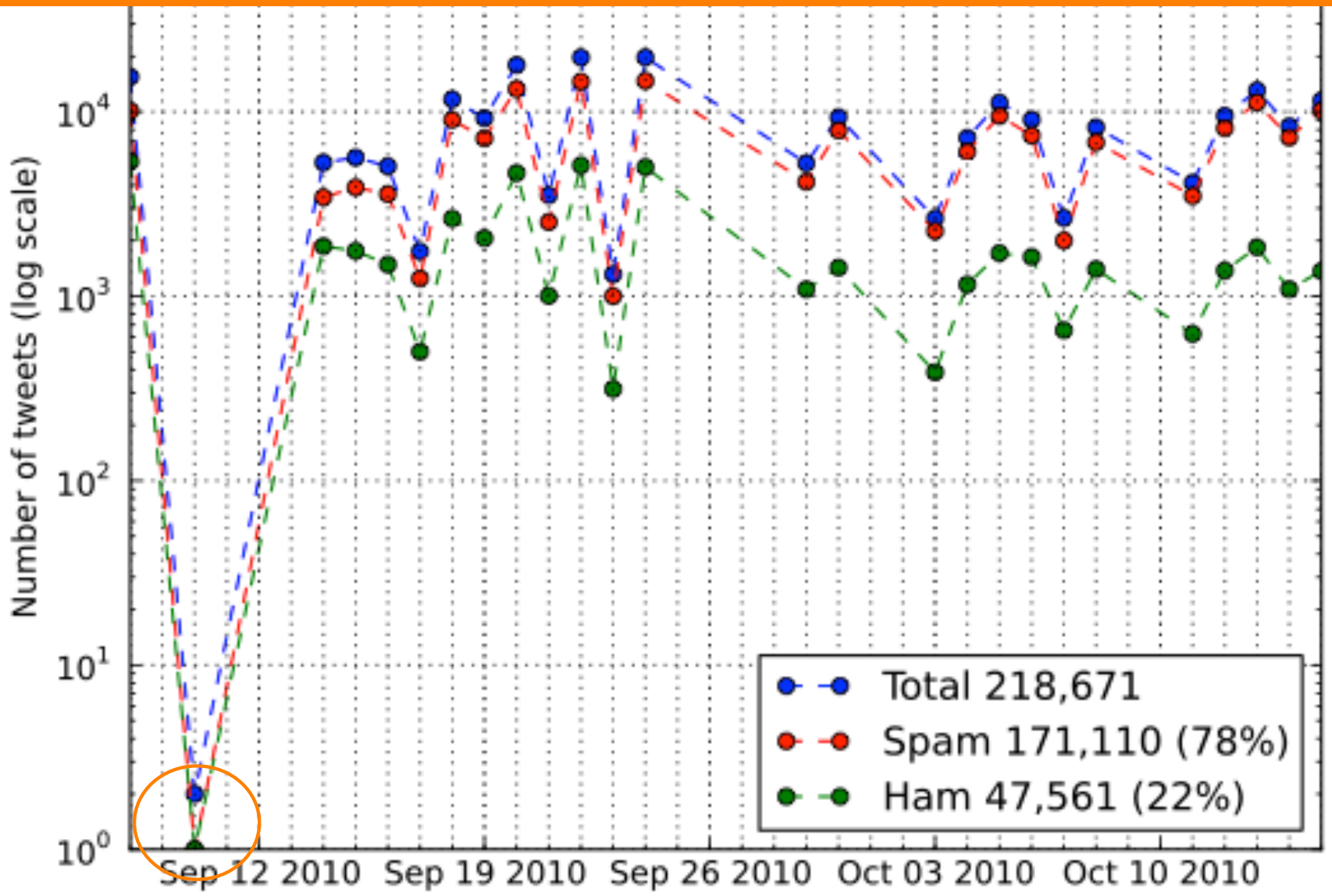
Particularly meaningless to connect categorical points with lines. ("Instances" likely hugely overcounts polymorphic malware)


Plots only 7 (x,y) points … which are discretized.
Wasted Y-axis real estate.
Not clear why 8-hour bins are appropriate.
Quantity of Responses
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
10
20
30
40
50
60
70
Response Delay (Hours)

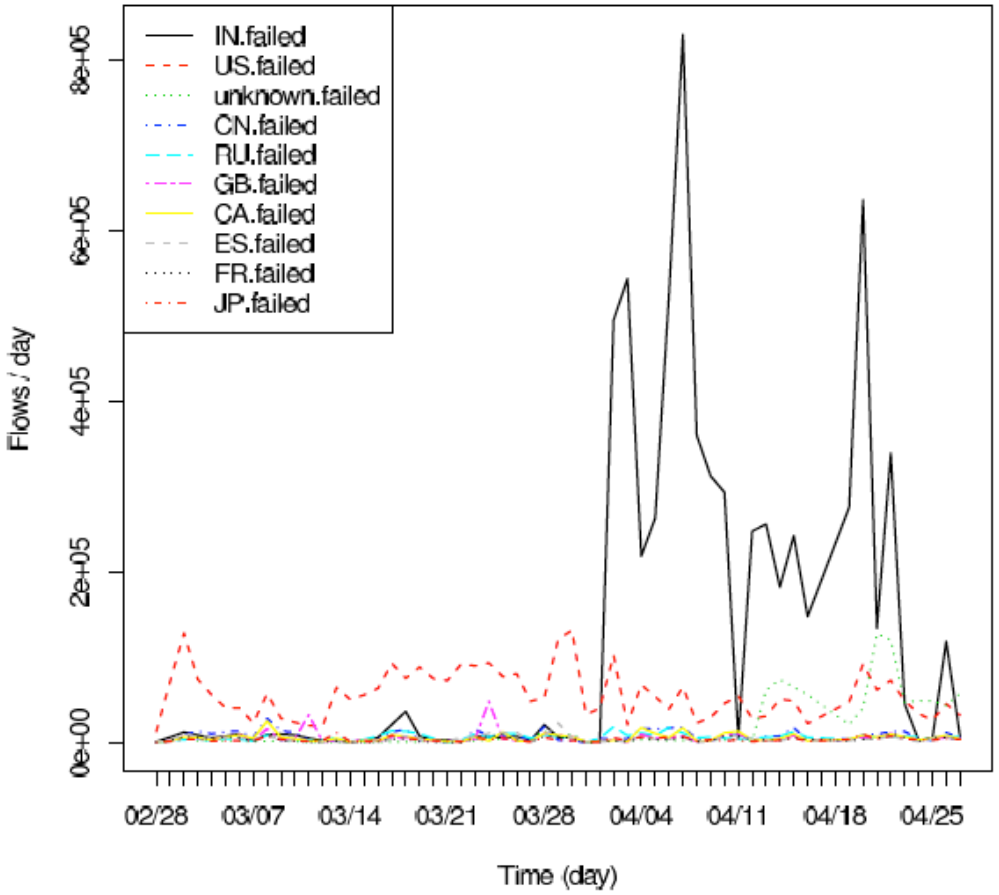


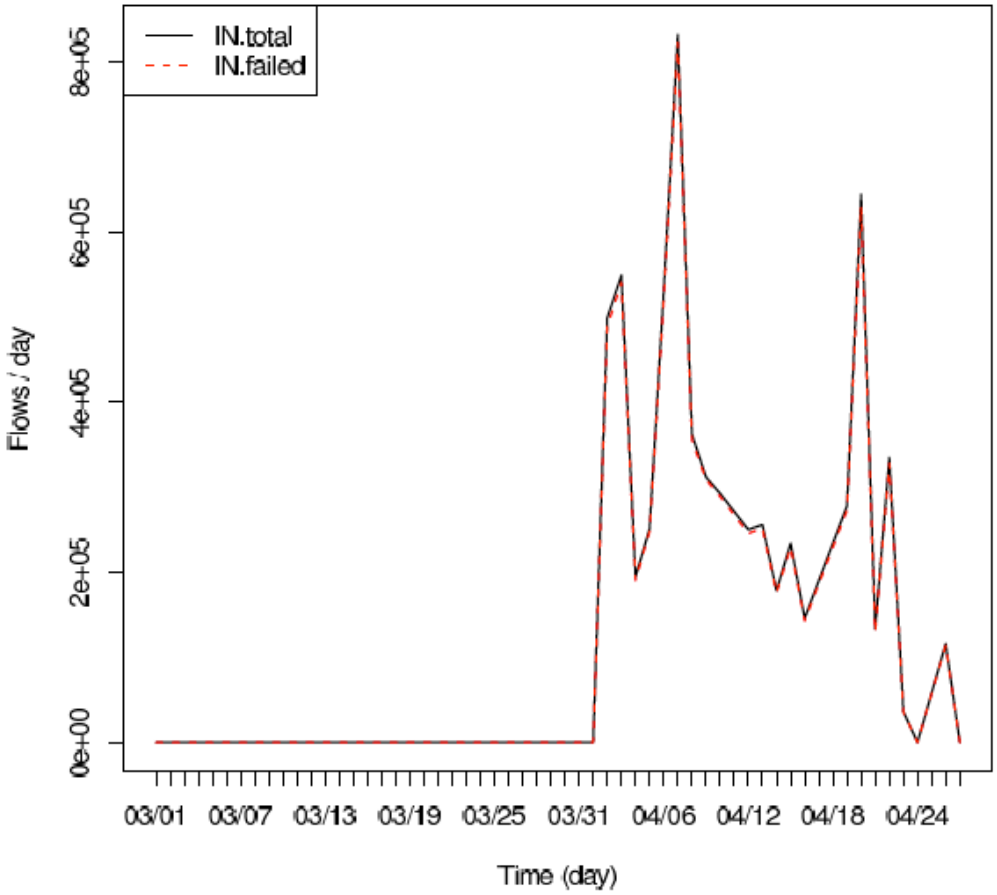
Hugely misleading compressed X-axis

Data glitch in second point visually dominates presentation. Straight-line interpolation on log-linear plot can be highly misleading.



Left-hand plot completely dominated by IN.failed.
Right-hand plot just shows that all of IN.total was IN.failed.





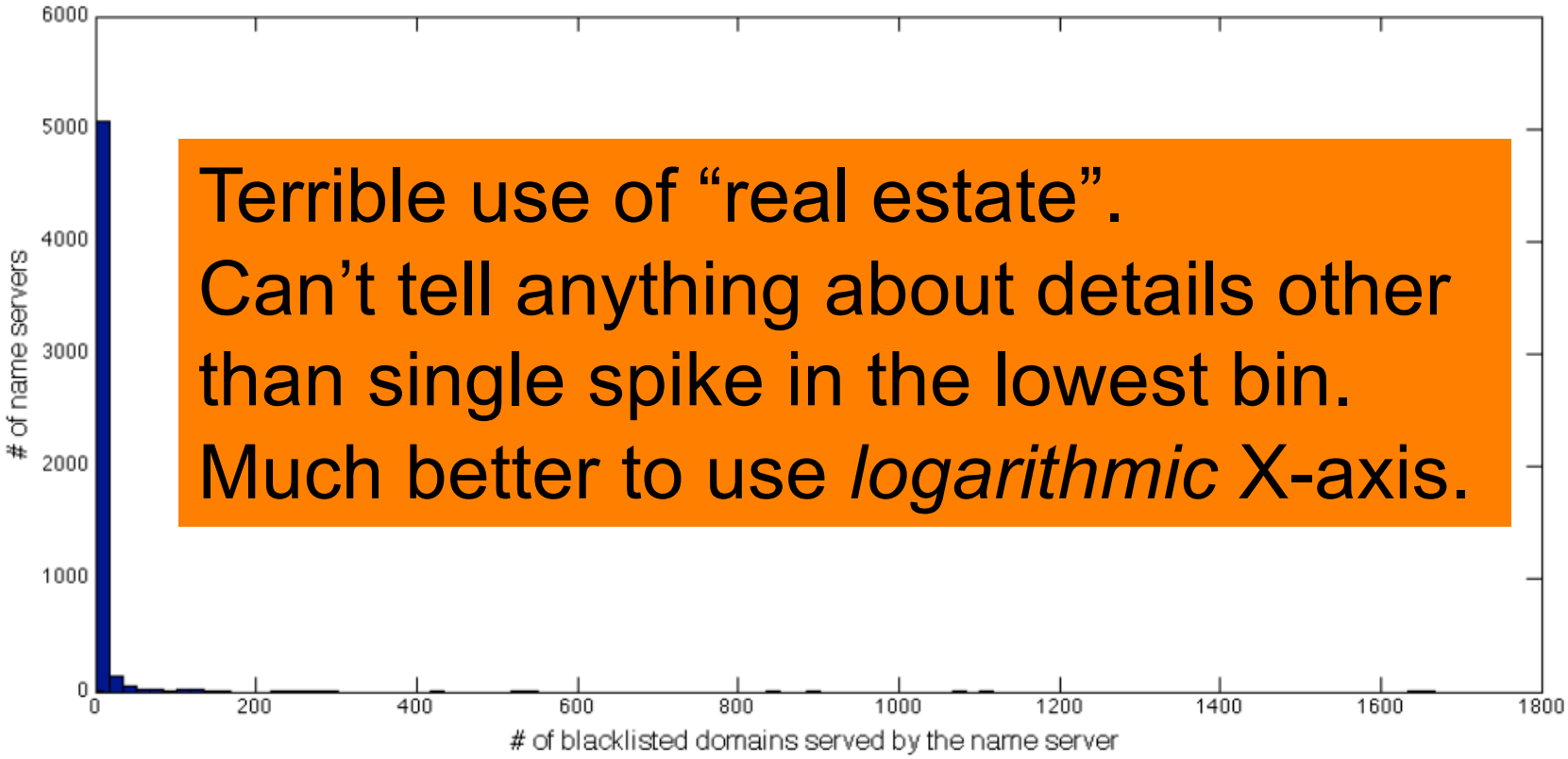
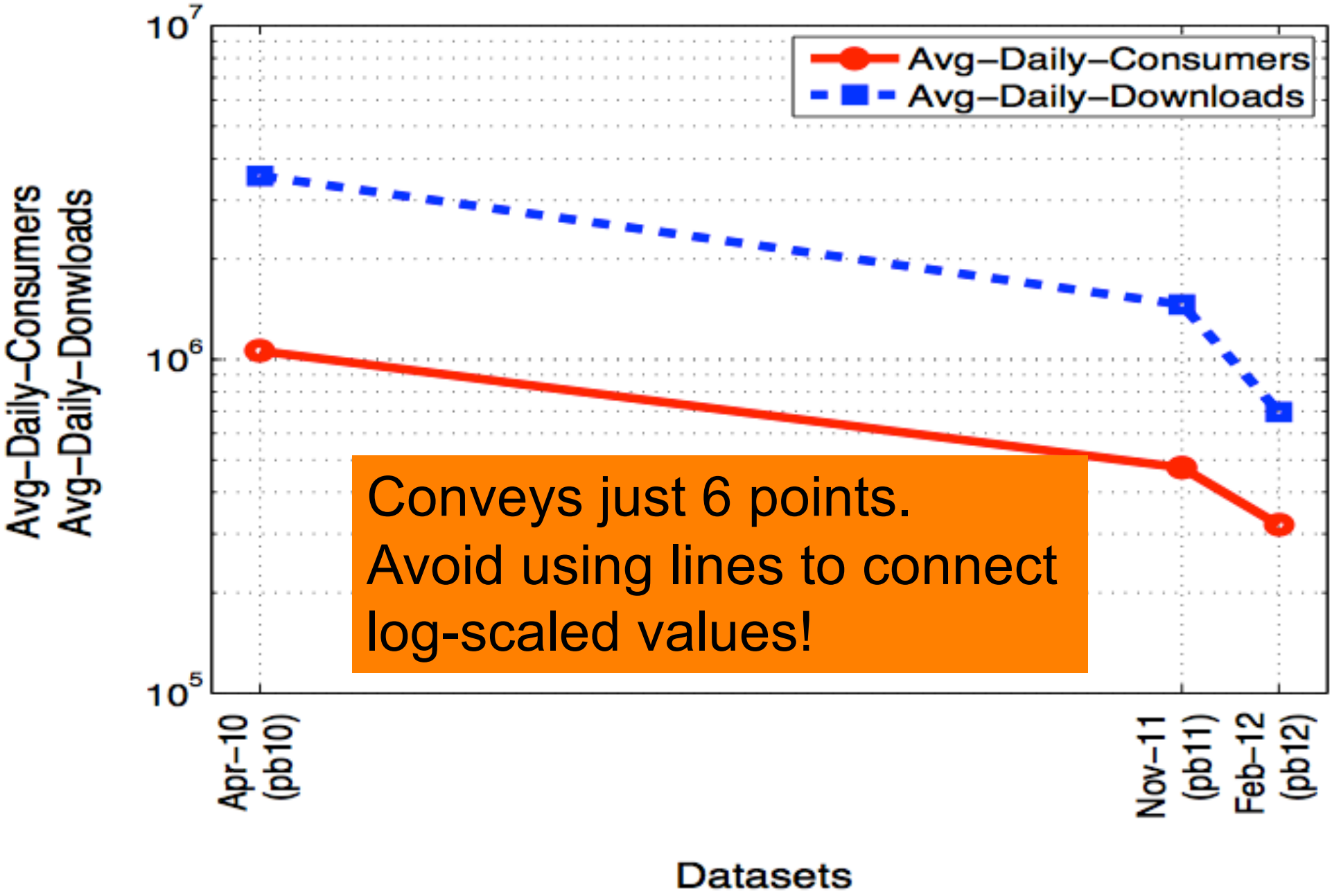


Figure 1: Histogram: Given an IP address, how many blacklisted domains use that IP address as a name server?


Avg–Daily–Consumers
Avg–Daily–Downloads
10^7
10^6
10^5
Avg–Daily–Consumers
Avg–Daily–Donwloads
Apr–10 (pb10)
Nov–11 (pb11)
Feb–12 (pb12)
Datasets
Conveys just 6 points.
Avoid using lines to connect log-scaled values!

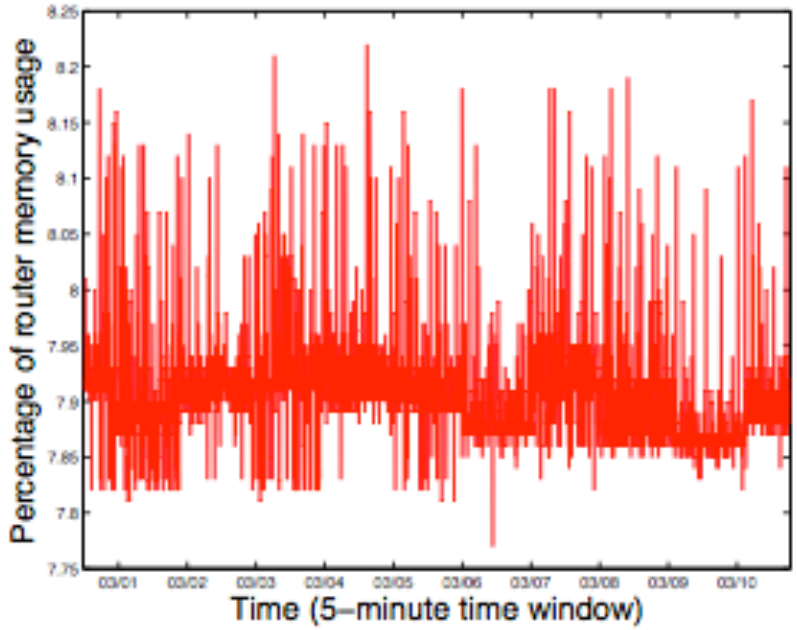
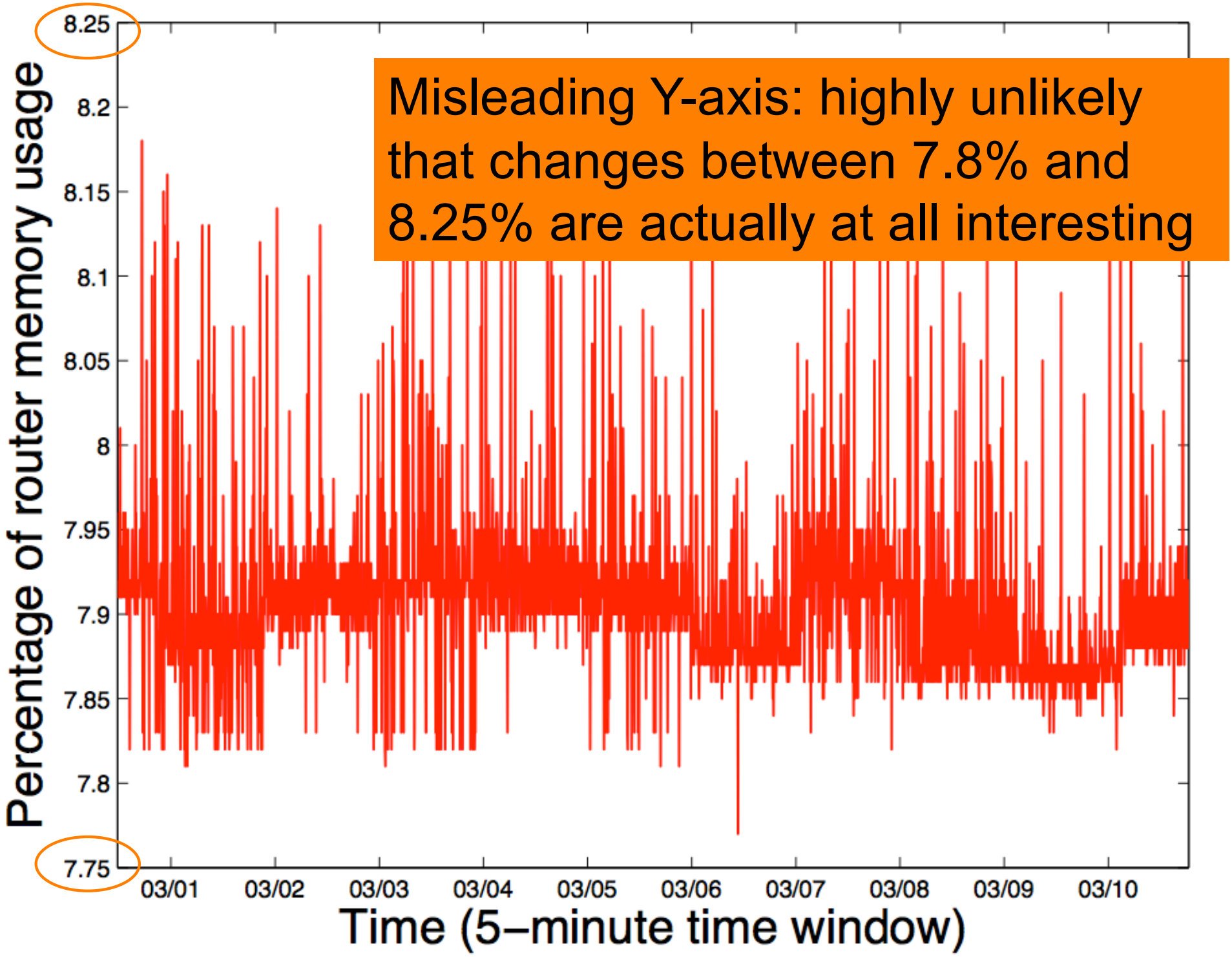


Figure 6: Total memory usage of traffic monitoring system.


Misleading Y-axis: highly unlikely that changes between 7.8% and 8.25% are actually at all interesting
Percentage of router memory usage
8.25
8.2
8.15
8.1
8.05
8
7.95
7.9
7.85
7.8
7.75
03/01
03/02
03/03
03/04
03/05
03/06
03/07
03/08
03/09
03/10
Time (5–minute time window)

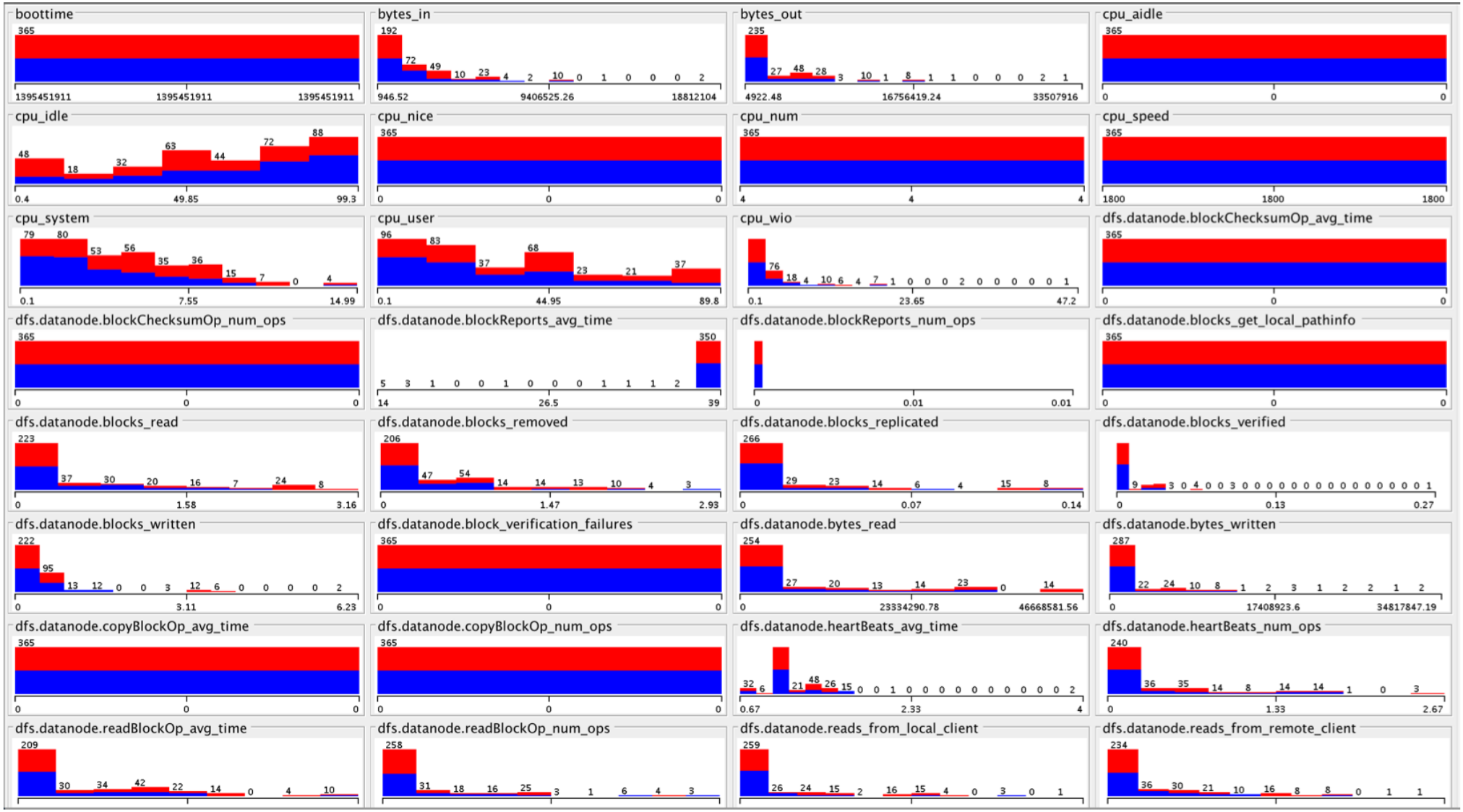
Trace	% Speculatively executed tasks
FB2009	1.22
FB2010	2.04
CC_b	1.01
CC_e	1.4

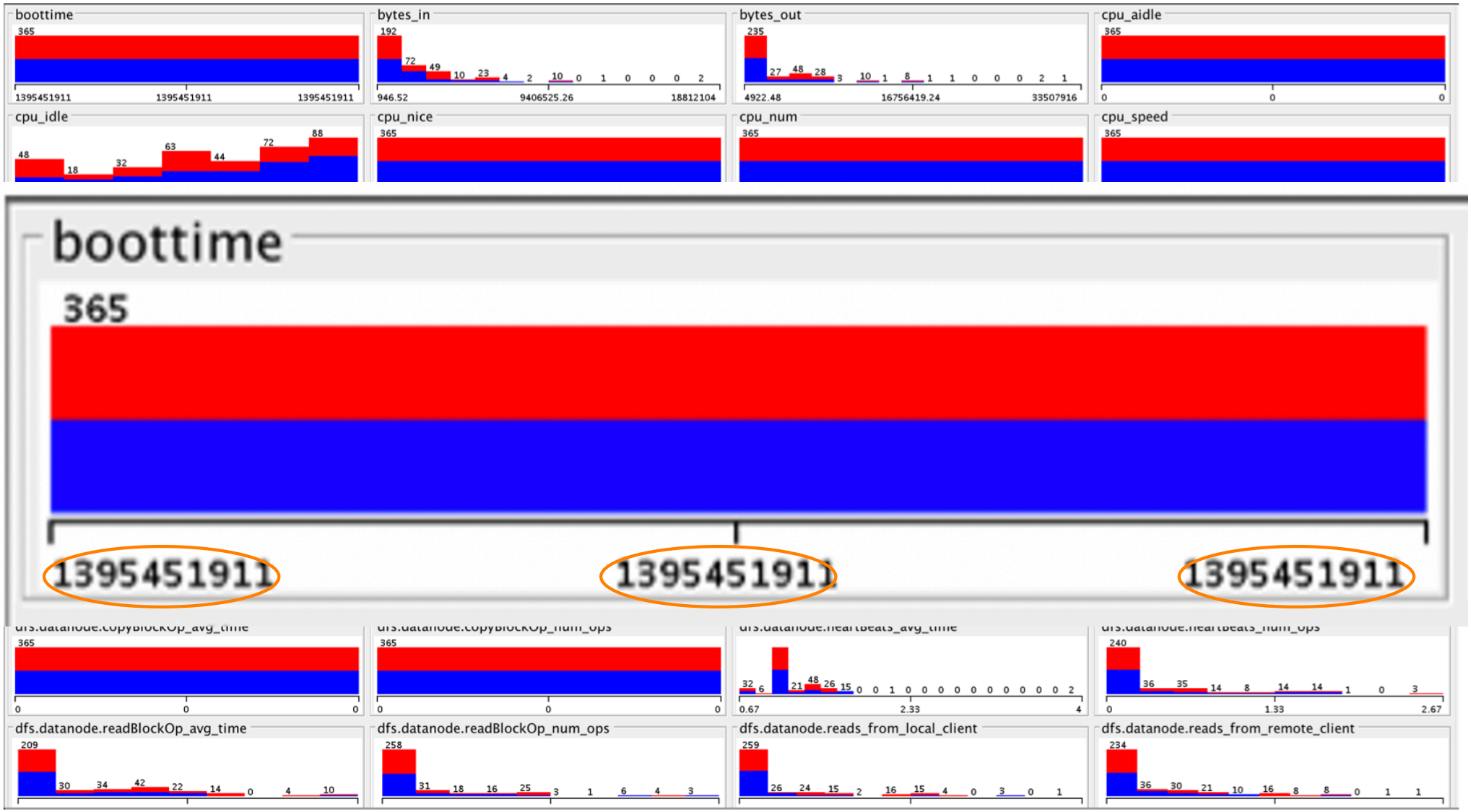
Trace	% of tasks that straggled even when they executed locally
FB2009	26.4
FB2010	39.2
CC_b	55
CC_e	56

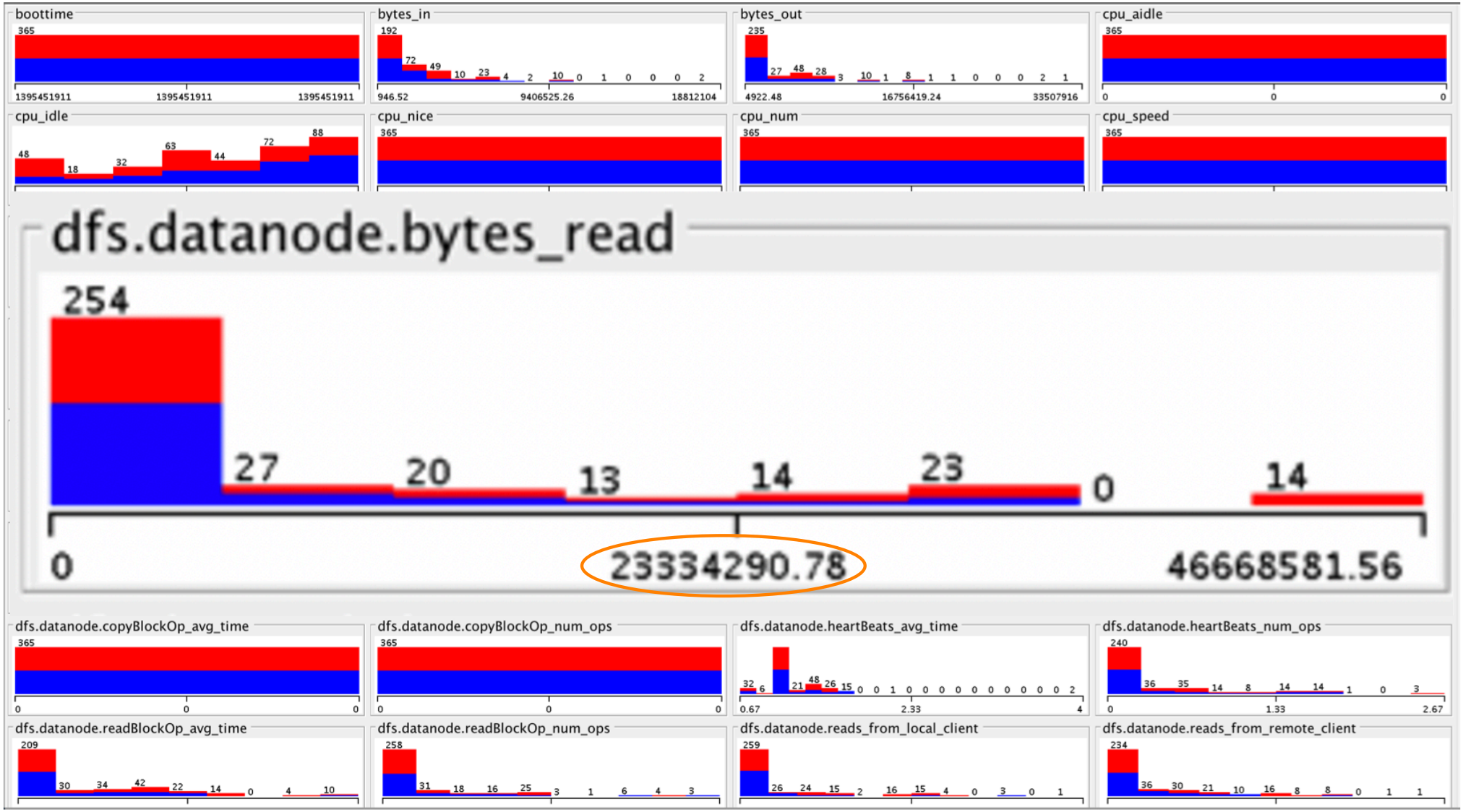
Trace	% of speculatively executed tasks that were killed
FB2009	77.9
FB2010	88.6
CC_b	74.4
CC_e	48.8

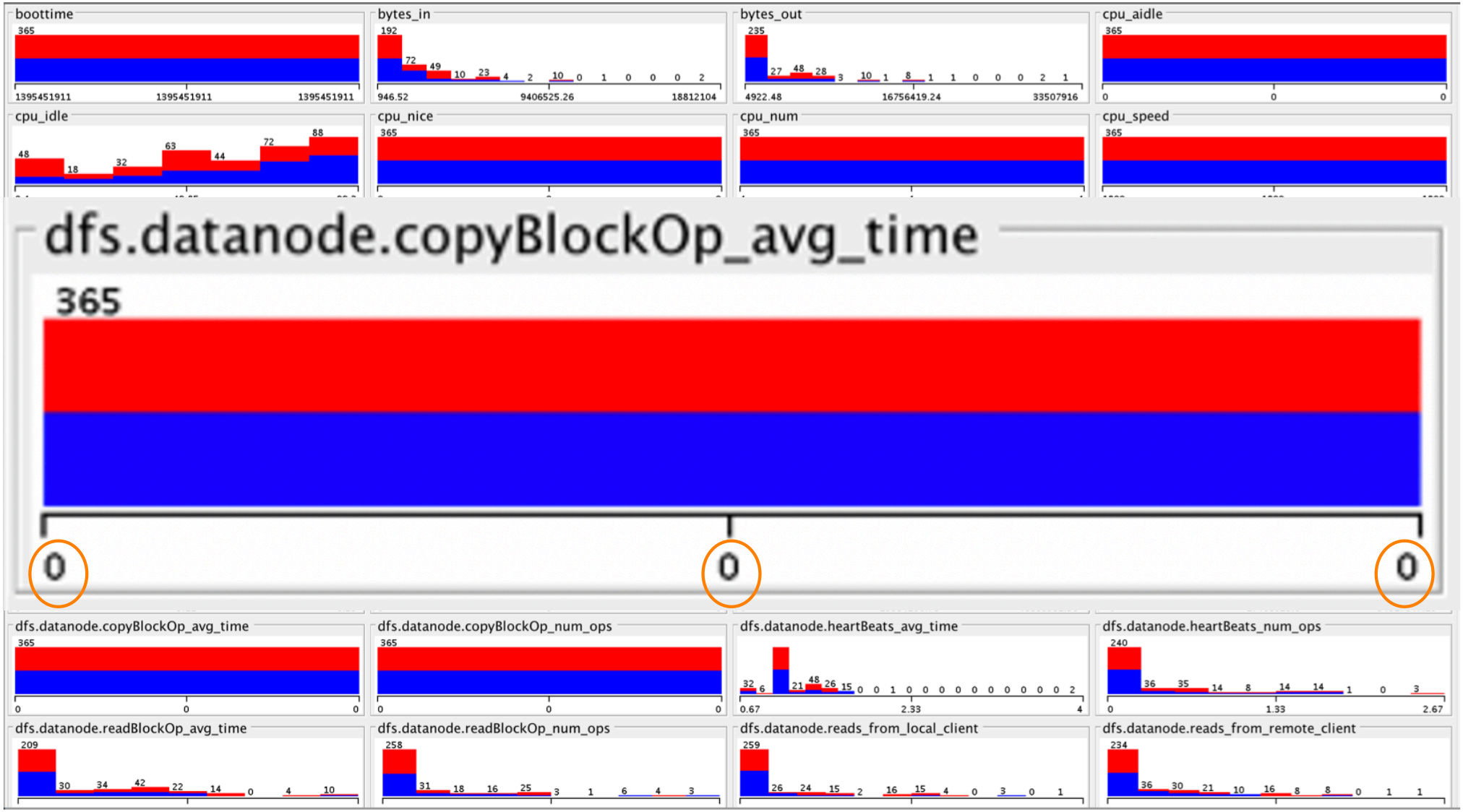
Trace	e killed
FB2009	57.57
FB2010	87.12
CC_b	97.4
CC_e	83.96

Large horizontal gaps make it visually a pain to read
Why 4 tables and not one table with 5 columns?

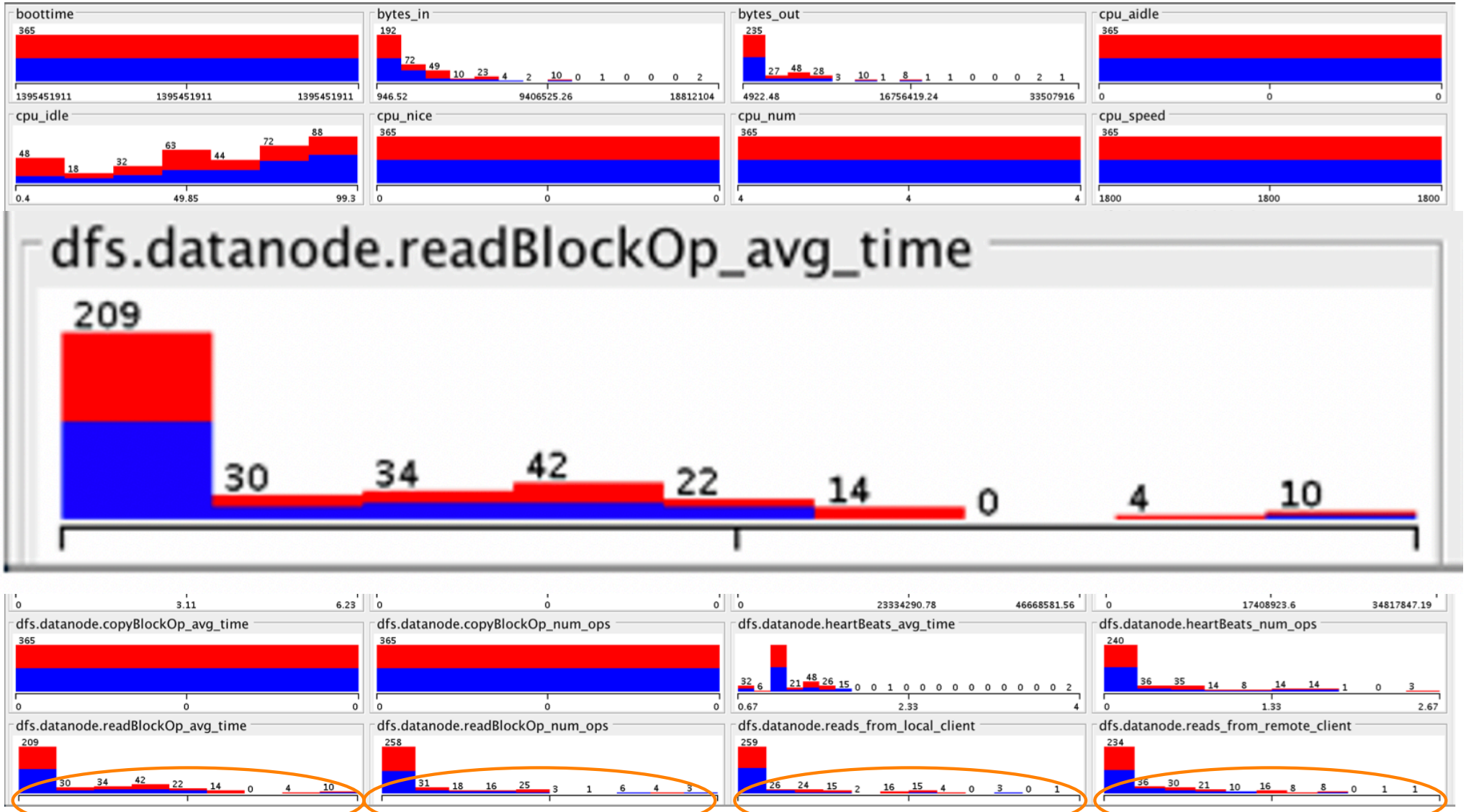

boottime
365
1395451911 1395451911 1395451911
bytes_in
192 72 49 10 23 4 2 10 0 1 0 0 0 2
946.52 9406525.26 18812104
bytes_out
235 27 48 28 3 10 1 8 1 1 0 0 0 2 1
4922.48 16756419.24 33507916
cpu_aidle
365
0 0 0
cpu_idle
48 18 32 63 44 72 88
0.4 49.85 99.3
cpu_nice
365
0 0 0
cpu_num
365
4 4 4
cpu_speed
365
1800 1800 1800
cpu_system
79 80 53 56 35 36 15 7 0 4
0.1 7.55 14.99
cpu_user
96 83 37 68 23 21 37
0.1 44.95 89.8
cpu_wio
76 18 4 10 6 4 7 1 0 0 0 2 0 0 0 0 0 1
0.1 23.65 47.2
dfs.datanode.blockChecksumOp_avg_time
365
0 0 0
dfs.datanode.blockChecksumOp_num_ops
365
0 0 0
dfs.datanode.blockReports_avg_time
350
5 3 1 0 0 1 0 0 0 1 1 1 2
14 26.5 39
dfs.datanode.blockReports_num_ops
0 0.01 0.01
dfs.datanode.blocks_get_local_pathinfo
365
0 0 0
dfs.datanode.blocks_read
223 37 30 20 16 7 24 8
0 1.58 3.16
dfs.datanode.blocks_removed
206 47 54 14 14 13 10 4 3
0 1.47 2.93
dfs.datanode.blocks_replicated
266 29 23 14 6 4 15 8
0 0.07 0.14
dfs.datanode.blocks_verified
9 3 0 4 0 0 3 0 0 0 0 0 0 0 0 0 0 0 0 0 1
0 0.13 0.27
dfs.datanode.blocks_written
222 95 13 12 0 0 3 12 6 0 0 0 0 2
0 3.11 6.23
dfs.datanode.block_verification_failures
365
0 0 0
dfs.datanode.bytes_read
254 27 20 13 14 23 0 14
0 23334290.78 46668581.56
dfs.datanode.bytes_written
287 22 24 10 8 1 2 3 1 2 2 1 2
0 17408923.6 34817847.19
dfs.datanode.copyBlockOp_avg_time
365
0 0 0
dfs.datanode.copyBlockOp_num_ops
365
0 0 0
dfs.datanode.heartBeats_avg_time
32 6 21 48 26 15 0 0 1 0 0 0 0 0 0 0 0 0 2
0.67 2.33 4
dfs.datanode.heartBeats_num_ops
240 36 35 14 8 14 14 1 0 3
0 1.33 2.67
dfs.datanode.readBlockOp_avg_time
209 30 34 42 22 14 0 4 10
dfs.datanode.readBlockOp_num_ops
258 31 18 16 25 3 1 6 4 3
dfs.datanode.reads_from_local_client
259 26 24 15 2 16 15 4 0 3 0 1
dfs.datanode.reads_from_remote_client
234 36 30 21 10 16 8 8 0 1 1

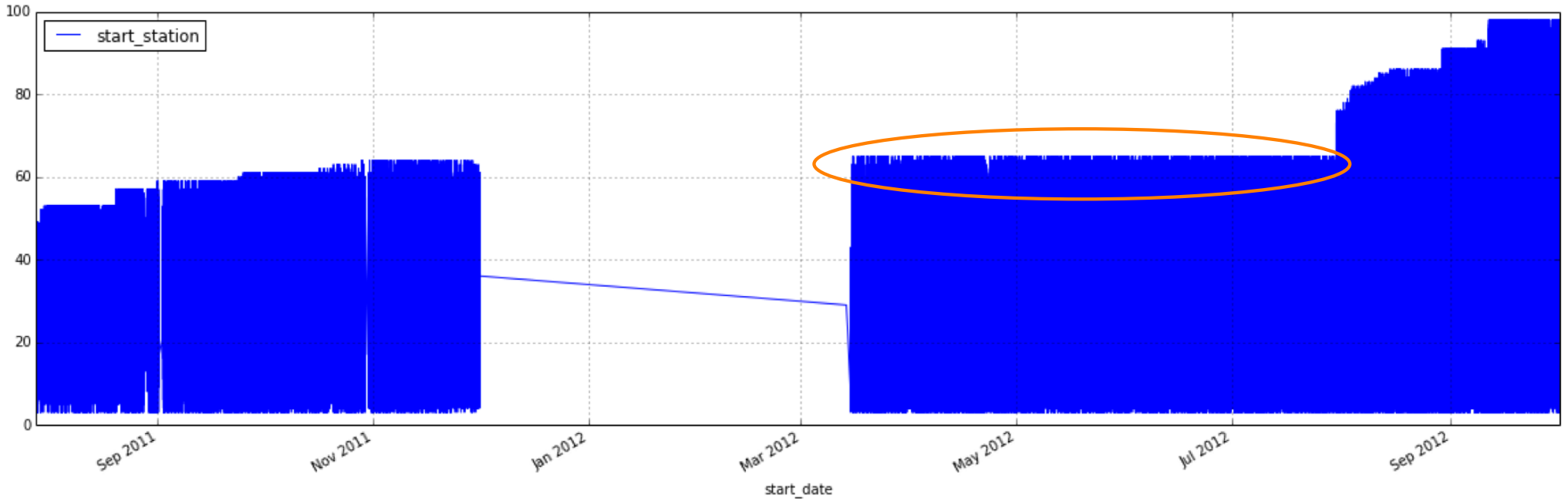

boottime
365
1395451911
1395451911
1395451911
bytes_in
192
72
49
10
23
4
2
10
0
1
0
0
0
2
946.52
9406525.26
18812104
bytes_out
235
27
48
28
3
10
1
8
1
1
0
0
0
2
1
4922.48
16756419.24
33507916
cpu_aidle
365
0
0
0
cpu_idle
48
18
32
63
44
72
88
cpu_nice
365
cpu_num
365
cpu_speed
365
boottime
365
1395451911
1395451911
1395451911
365
0
0
0
365
0
0
0
0.67
2.33
4
240
36
35
14
8
14
14
1
0
3
0
1.33
2.67
dfs.datanode.readBlockOp_avg_time
209
30
34
42
22
14
0
4
10
dfs.datanode.readBlockOp_num_ops
258
31
18
16
25
3
1
6
4
3
dfs.datanode.reads_from_local_client
259
26
24
15
2
16
15
4
0
3
0
1
dfs.datanode.reads_from_remote_client
234
36
30
21
10
16
8
8
0
1
1


boottime
365
1395451911
1395451911
1395451911
bytes_in
192
72
49
10
23
4
2
10
0
1
0
0
0
2
946.52
9406525.26
18812104
bytes_out
235
27
48
28
3
10
1
8
1
1
0
0
0
2
1
4922.48
16756419.24
33507916
cpu_aidle
365
0
0
0
cpu_idle
48
18
32
63
44
72
88
cpu_nice
365
cpu_num
365
cpu_speed
365
dfs.datanode.bytes_read
254
27
20
13
14
23
0
14
0
23334290.78
46668581.56
dfs.datanode.copyBlockOp_avg_time
365
0
0
0
dfs.datanode.copyBlockOp_num_ops
365
0
0
0
dfs.datanode.heartBeats_avg_time
32
6
21
48
26
15
0
0
1
0
0
0
0
0
0
0
0
2
0.67
2.33
4
dfs.datanode.heartBeats_num_ops
240
36
35
14
8
14
14
1
0
3
0
1.33
2.67
dfs.datanode.readBlockOp_avg_time
209
30
34
42
22
14
0
4
10
dfs.datanode.readBlockOp_num_ops
258
31
18
16
25
3
1
6
4
3
dfs.datanode.reads_from_local_client
259
26
24
15
2
16
15
4
0
3
0
1
dfs.datanode.reads_from_remote_client
234
36
30
21
10
16
8
8
0
1
1

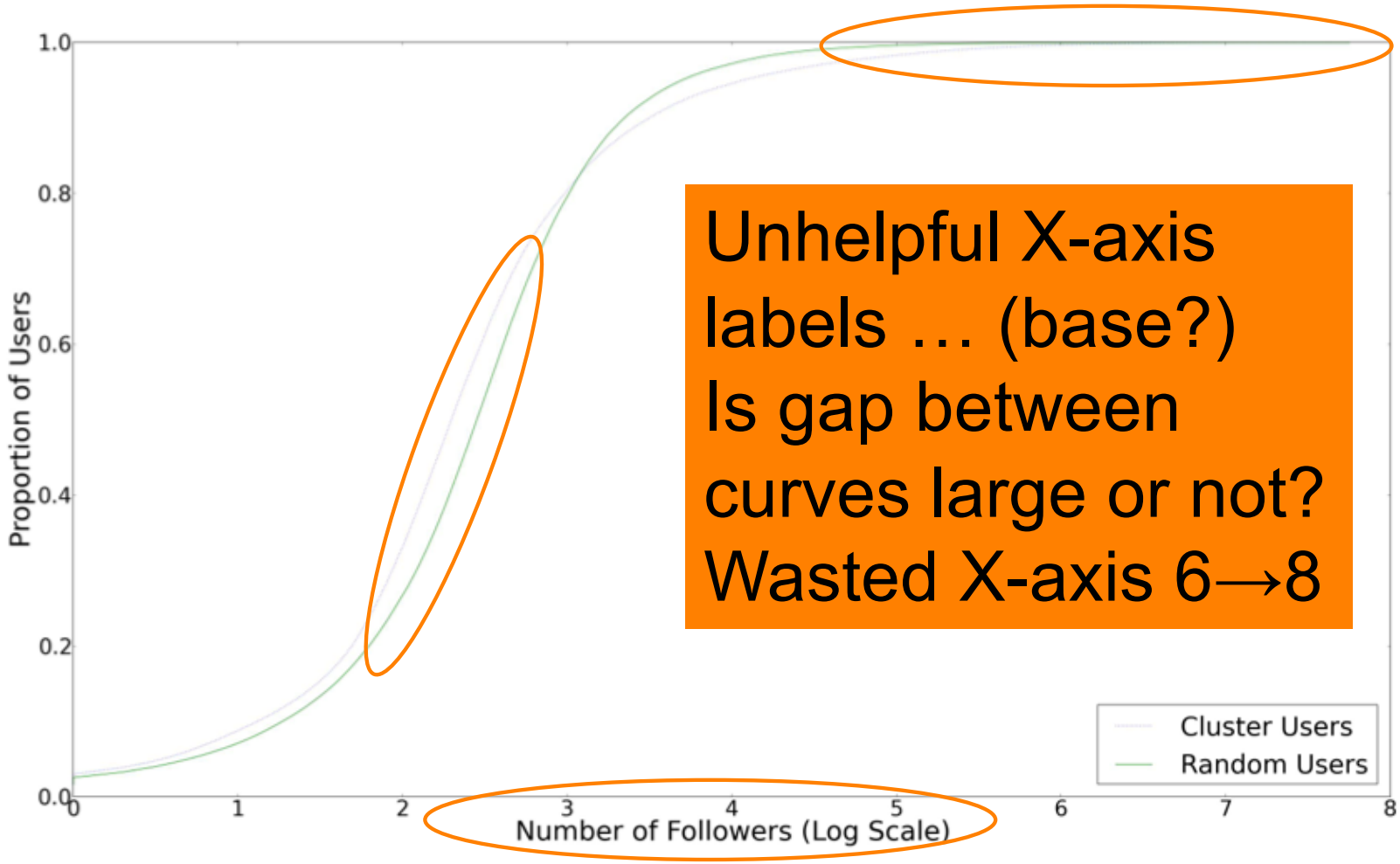

boottime
365
1395451911
1395451911
1395451911
bytes_in
192
72
49
10
23
4
2
10
0
1
0
0
0
2
946.52
9406525.26
18812104
bytes_out
235
27
48
28
3
10
1
8
1
1
0
0
0
2
1
4922.48
16756419.24
33507916
cpu_aidle
365
0
0
0
cpu_idle
48
18
32
63
44
72
88
cpu_nice
365
cpu_num
365
cpu_speed
365
dfs.datanode.copyBlockOp_avg_time
365
0
0
0
dfs.datanode.copyBlockOp_avg_time
365
0
0
0
dfs.datanode.copyBlockOp_num_ops
365
0
0
0
dfs.datanode.heartBeats_avg_time
32
6
21
48
26
15
0
0
1
0
0
0
0
0
0
0
0
2
0.67
2.33
4
dfs.datanode.heartBeats_num_ops
240
36
35
14
8
14
14
1
0
3
0
1.33
2.67
dfs.datanode.readBlockOp_avg_time
209
30
34
42
22
14
0
4
10
dfs.datanode.readBlockOp_num_ops
258
31
18
16
25
3
1
6
4
3
dfs.datanode.reads_from_local_client
259
26
24
15
2
16
15
4
0
3
0
1
dfs.datanode.reads_from_remote_client
234
36
30
21
10
16
8
8
0
1
1


boottime
365
1395451911
1395451911
1395451911
bytes_in
192
72
49
10
23
4
2
10
0
1
0
0
0
2
946.52
9406525.26
18812104
bytes_out
235
27
48
28
3
10
1
8
1
1
0
0
0
2
1
4922.48
16756419.24
33507916
cpu_aidle
365
0
0
0
cpu_idle
48
18
32
63
44
72
88
cpu_nice
365
cpu_num
365
cpu_speed
365
dfs.datanode.blockReports_num_ops
0
0.01
0.01
dfs.datanode.copyBlockOp_avg_time
365
0
0
0
dfs.datanode.copyBlockOp_num_ops
365
0
0
0
dfs.datanode.heartBeats_avg_time
32
6
21
48
26
15
0
0
1
0
0
0
0
0
0
0
0
2
0.67
2.33
4
dfs.datanode.heartBeats_num_ops
240
36
35
14
8
14
14
1
0
3
0
1.33
2.67
dfs.datanode.readBlockOp_avg_time
209
30
34
42
22
14
0
4
10
dfs.datanode.readBlockOp_num_ops
258
31
18
16
25
3
1
6
4
3
dfs.datanode.reads_from_local_client
259
26
24
15
2
16
15
4
0
3
0
1
dfs.datanode.reads_from_remote_client
234
36
30
21
10
16
8
8
0
1
1


boottime
365
1395451911
1395451911
1395451911
bytes_in
192
72
49
10
23
4
2
10
0
1
0
0
0
2
946.52
9406525.26
18812104
bytes_out
235
27
48
28
3
10
1
8
1
1
0
0
0
2
1
4922.48
16756419.24
33507916
cpu_aidle
365
0
0
0
cpu_idle
48
18
32
63
44
72
88
0.4
49.85
99.3
cpu_nice
365
0
0
0
cpu_num
365
4
4
4
cpu_speed
365
1800
1800
1800
dfs.datanode.readBlockOp_avg_time
209
30
34
42
22
14
0
4
10
0
3.11
6.23
0
0
0
0
23334290.78
46668581.56
0
17408923.6
34817847.19
dfs.datanode.copyBlockOp_avg_time
365
0
0
0
dfs.datanode.copyBlockOp_num_ops
365
0
0
0
dfs.datanode.heartBeats_avg_time
32
6
21
48
26
15
0
0
1
0
0
0
0
0
0
0
0
2
0.67
2.33
4
dfs.datanode.heartBeats_num_ops
240
36
35
14
8
14
14
1
0
3
0
1.33
2.67
dfs.datanode.readBlockOp_avg_time
209
30
34
42
22
14
0
4
10
dfs.datanode.readBlockOp_num_ops
258
31
18
16
25
3
1
6
4
3
dfs.datanode.reads_from_local_client
259
26
24
15
2
16
15
4
0
3
0
1
dfs.datanode.reads_from_remote_client
234
36
30
21
10
16
8
8
0
1
1



Highly distracting central gap
Just what do the authors want us to take away from this?



(a) Number of Followers for Cluster Users and a Random Sample of Users

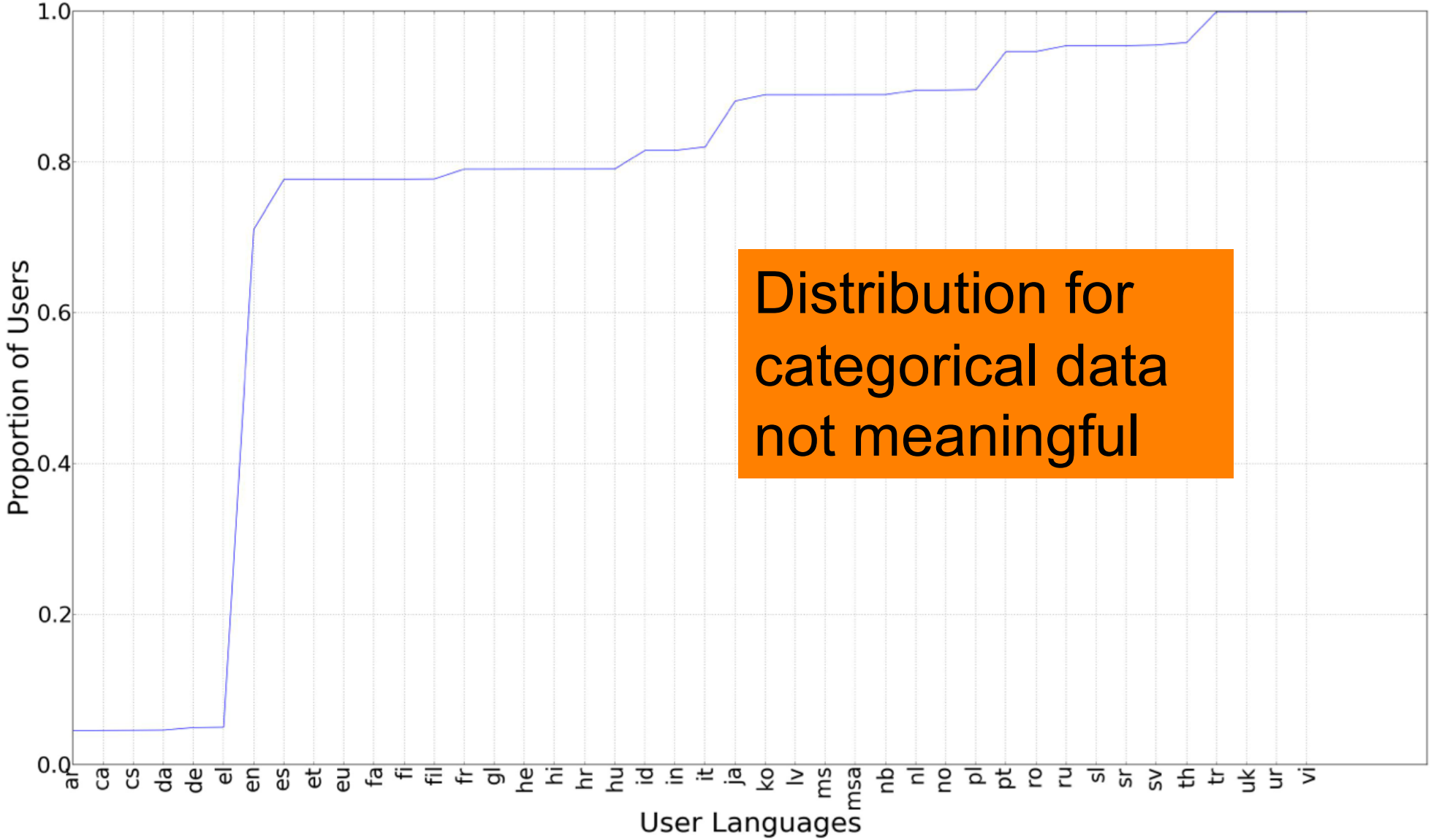
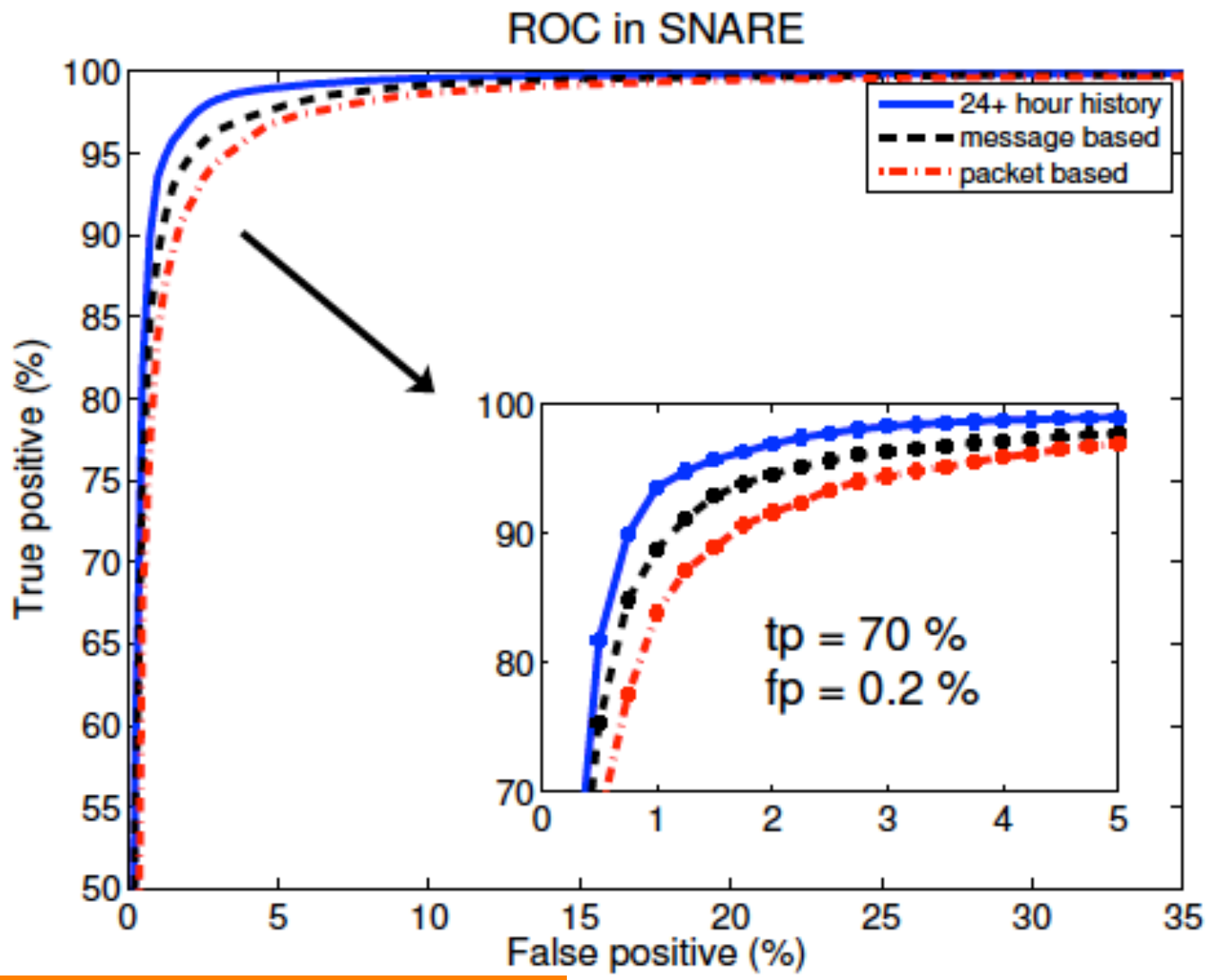


Fig. 7: CDF of user account languages for users in size 2 clusters



An example of good use of plot “real estate”