Internet Mail — Past, Present, and (a bit of) the Future

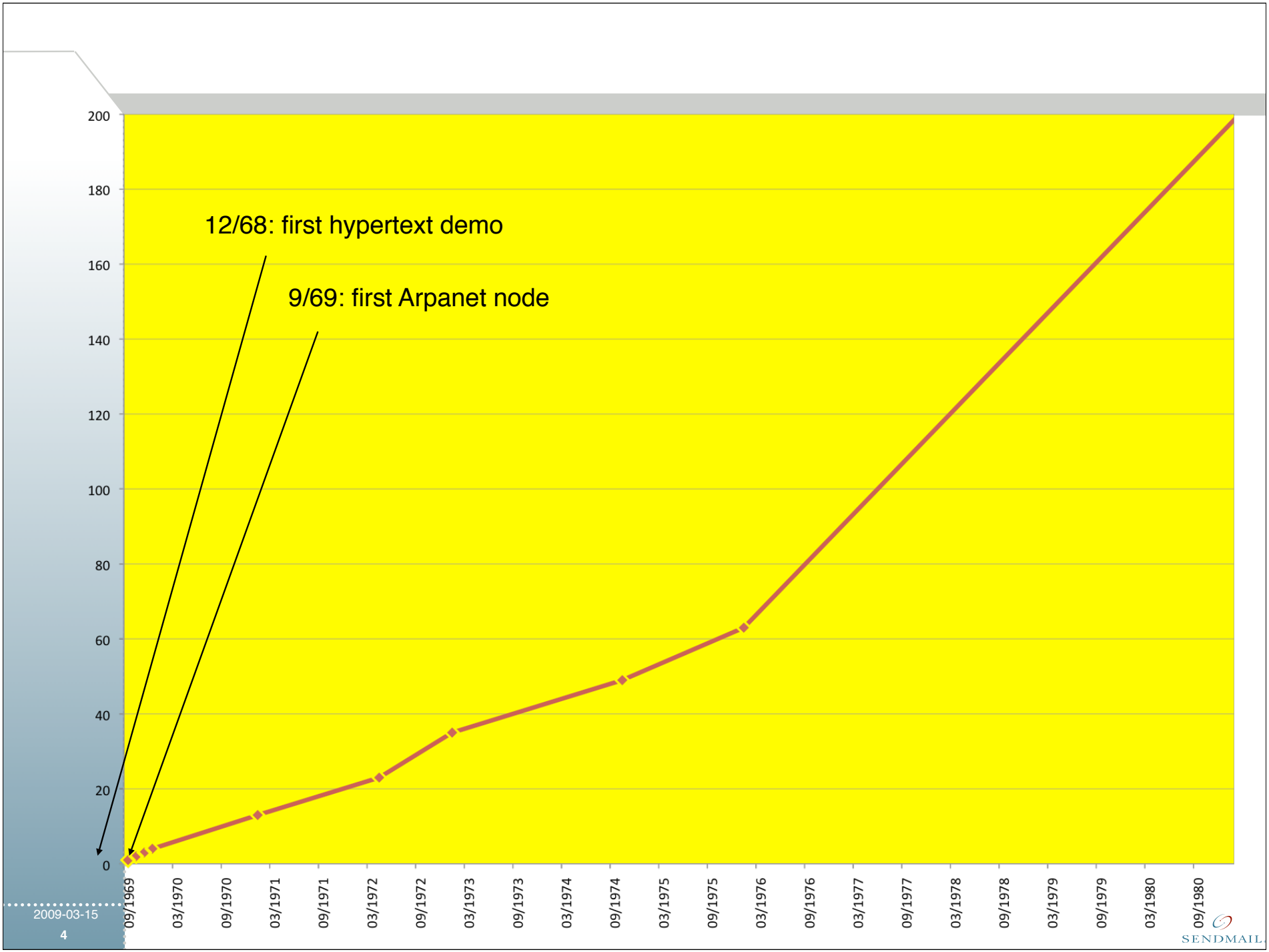
Eric Allman
Chief Science Officer
Sendmail, Inc.
March 2009

Introduction and Outline

- A history of email and related technologies from the very early days through present
- Some observations about spam, email security, authentication, and reputation
- Very fuzzy speculations about the future of email and messaging in general

Internet Pre-History
1968–1981

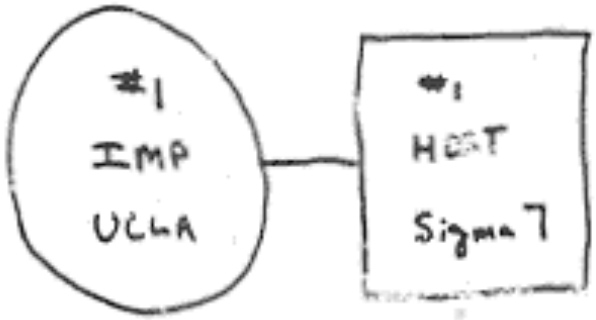
Arpanet: 0–213 hosts


200
180
160
140
120
100
80
60
40
20
0
12/68: first hypertext demo
9/69: first Arpanet node
09/1969
03/1970
09/1970
03/1971
09/1971
03/1972
09/1972
03/1973
09/1973
03/1974
09/1974
03/1975
09/1975
03/1976
09/1976
03/1977
09/1977
03/1978
09/1978
03/1979
09/1979
03/1980
09/1980

Genesis of the Arpanet

- Sprung into being September 2, 1969
- One host at UCLA
- No one to talk to and no place to go….

Arpanet September 1969



THE ARPA NETWORK

SEPT. 1969

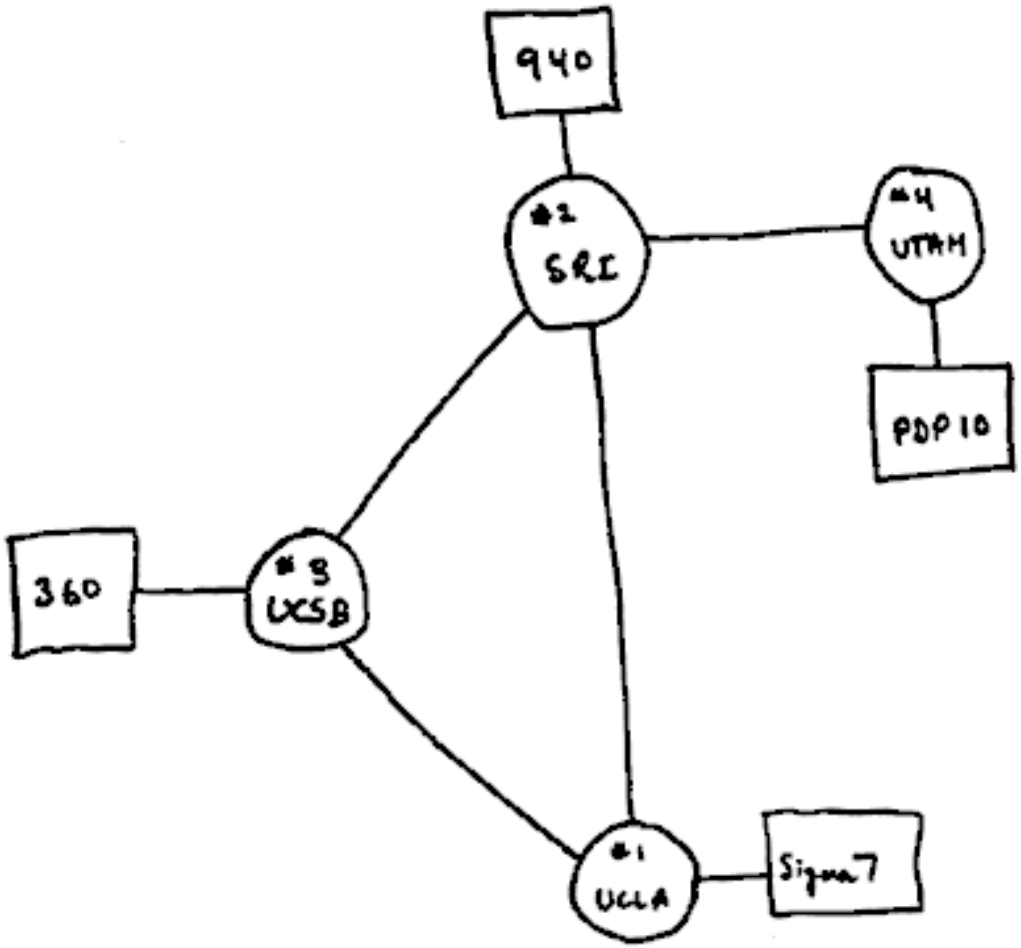
1 NODE

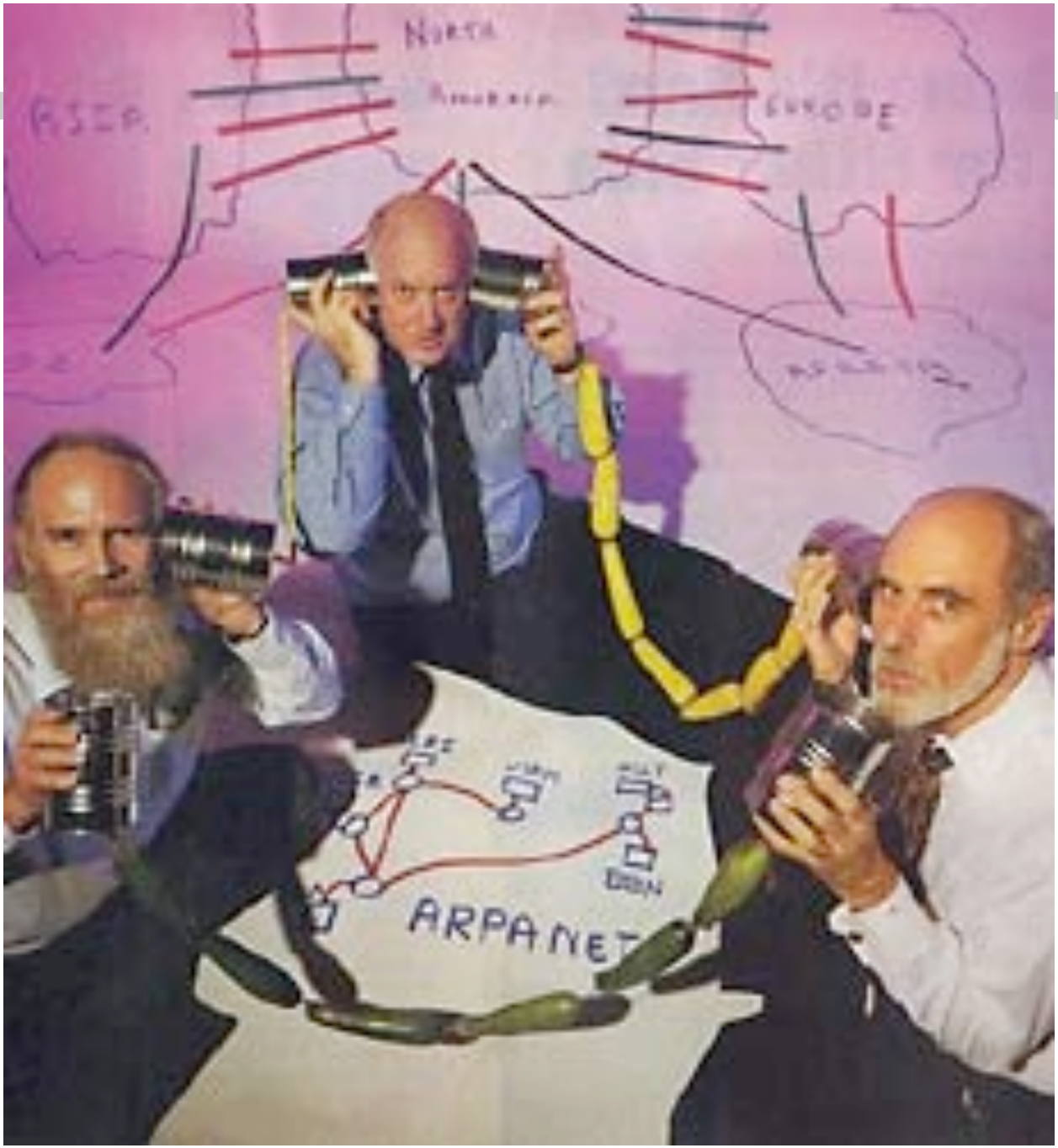
FIGURE 6.1 Drawing of September 1969
(Courtesy of Alex McKenzie)

Genesis of the Arpanet

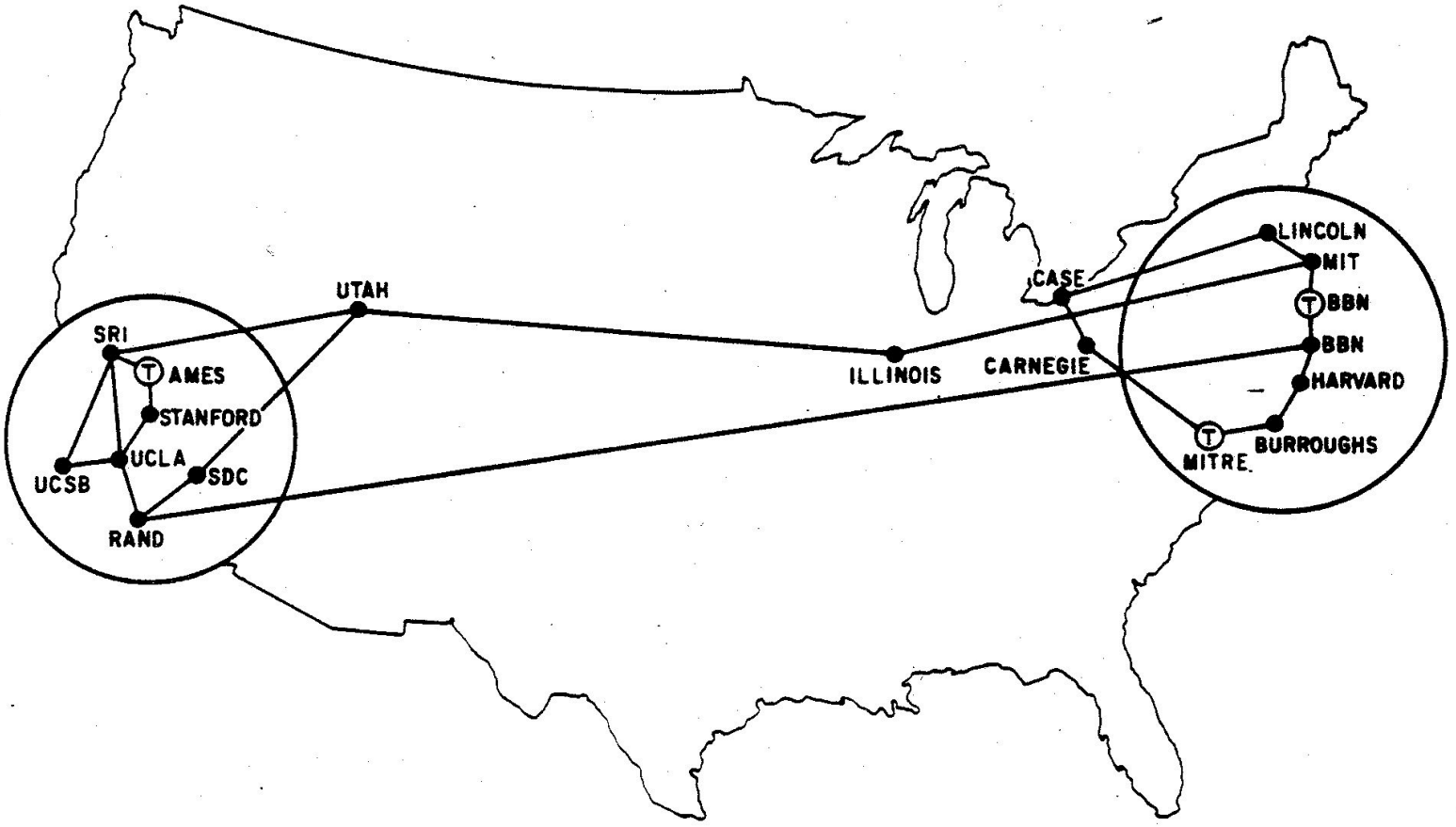
- Sprung into being September 2, 1969
- One host at UCLA
- No one to talk to and no place to go….
- Soon, hosts added at Stanford Research Institute, University of California Santa Barbara, and University of Utah (one per month)

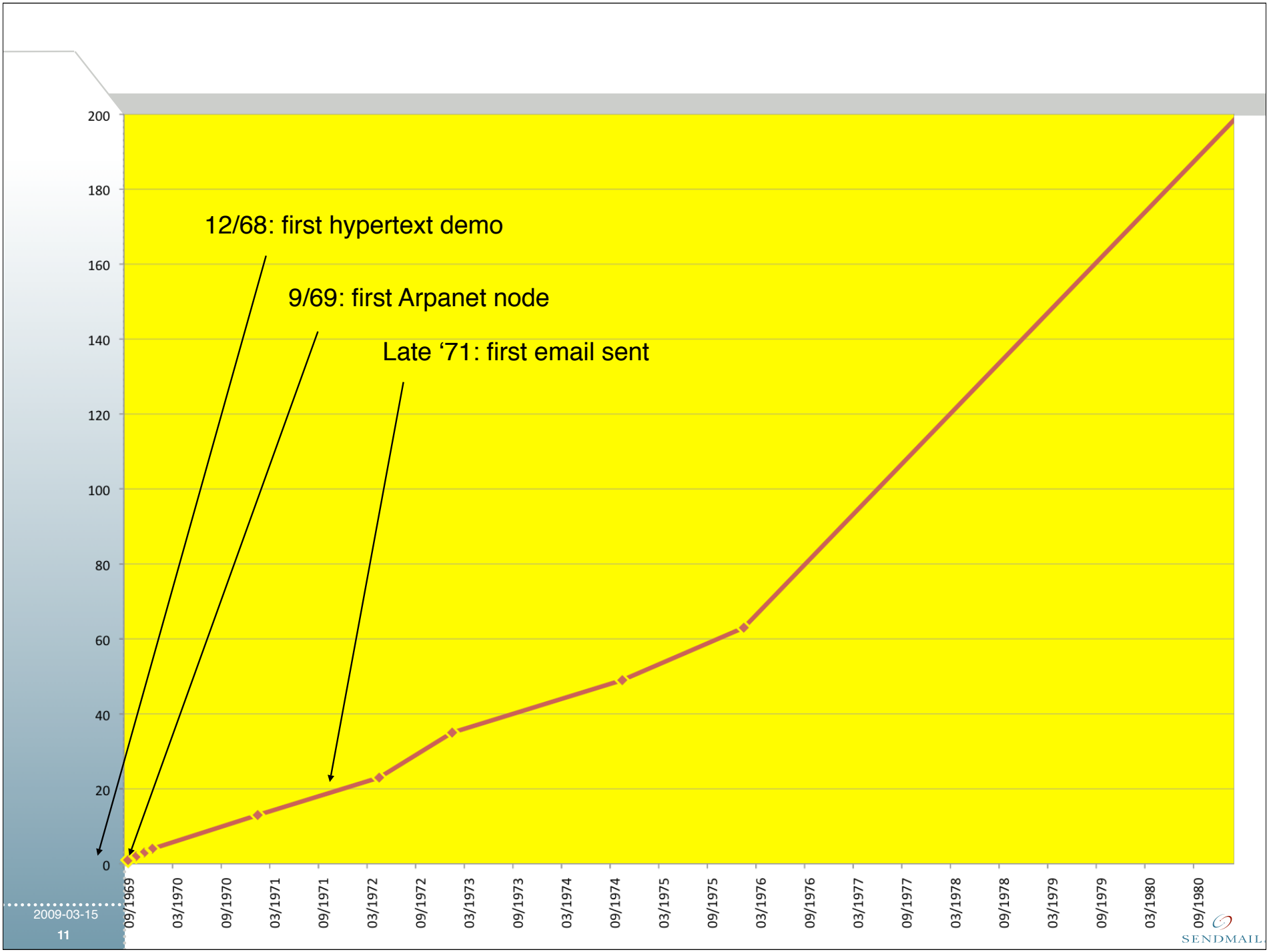
Arpanet Late 1969 (Logical View)




ARPANET

Arpanet September 1971



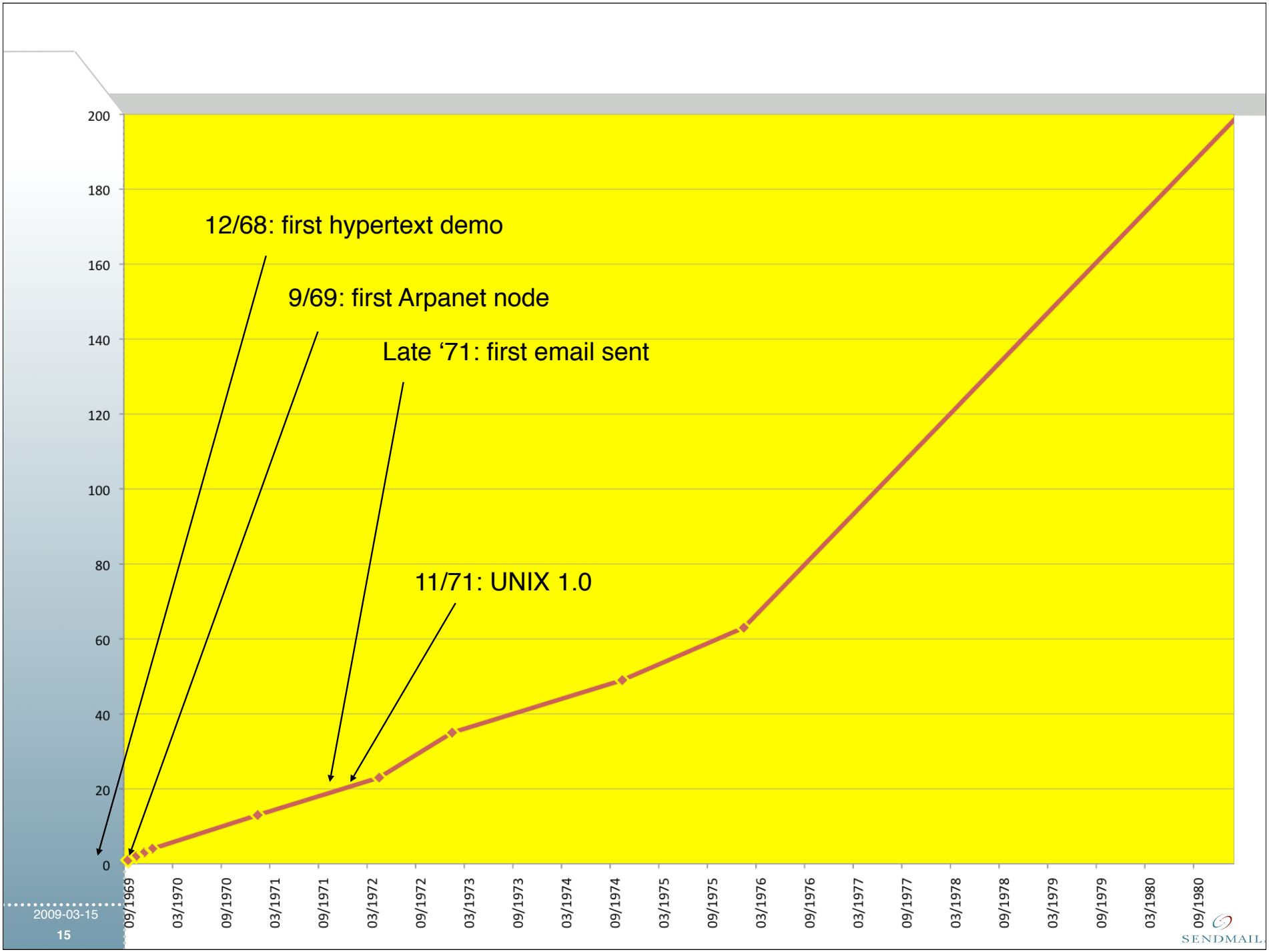

200
180
160
140
120
100
80
60
40
20
0
12/68: first hypertext demo
9/69: first Arpanet node
Late ‘71: first email sent
09/1969
03/1970
09/1970
03/1971
09/1971
03/1972
09/1972
03/1973
09/1973
03/1974
09/1974
03/1975
09/1975
03/1976
09/1976
03/1977
09/1977
03/1978
09/1978
03/1979
09/1979
03/1980
09/1980

Ray Tomlinson — the Real Father of Email

Ray Tomlinson

- Really is the Father of Email
- Bolt Beranek & Newman
- Email just append-only file transfer to a special file
- Late 1971: linked BBNA and BBNB

The First Two Email Hosts


200
180
160
140
120
100
80
60
40
20
0
12/68: first hypertext demo
9/69: first Arpanet node
Late '71: first email sent
11/71: UNIX 1.0
09/1969
03/1970
09/1970
03/1971
09/1971
03/1972
09/1972
03/1973
09/1973
03/1974
09/1974
03/1975
09/1975
03/1976
09/1976
03/1977
09/1977
03/1978
09/1978
03/1979
09/1979
03/1980
09/1980

Ken Thompson and Dennis Ritchie

1973

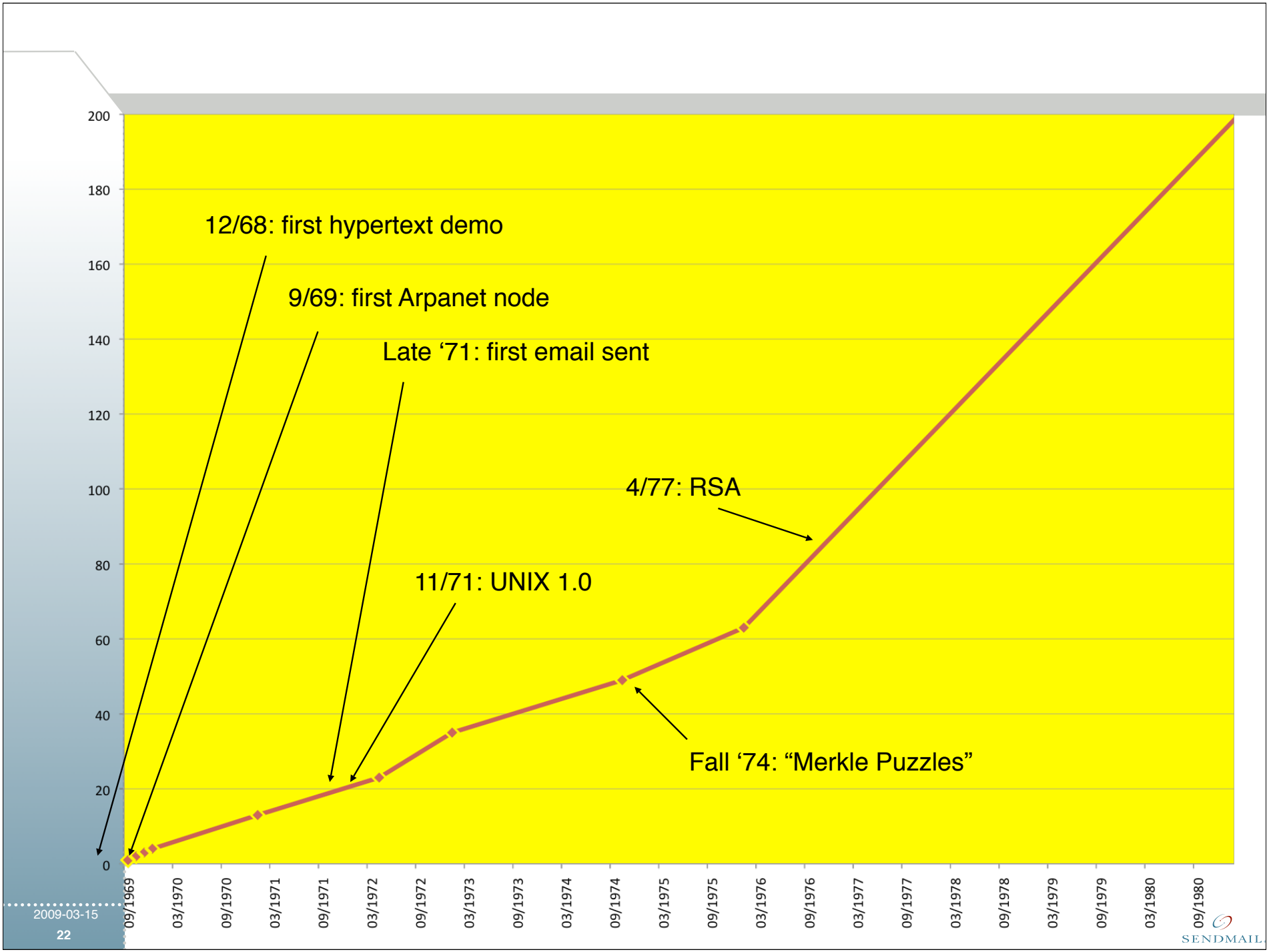
Beginning of a truly exciting time at Berkeley

Bill Joy

Kirk McKusick

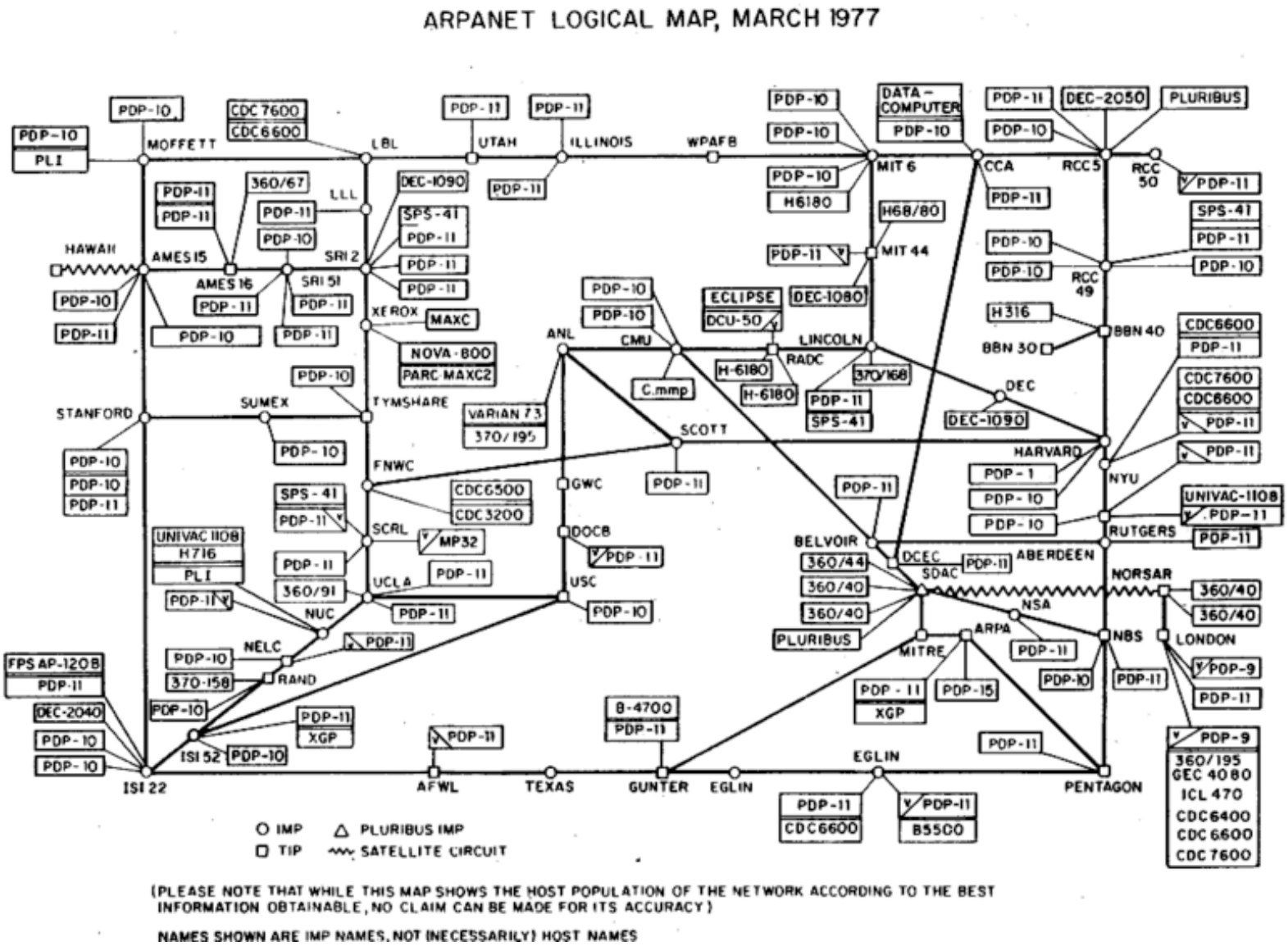
Eric Schmidt

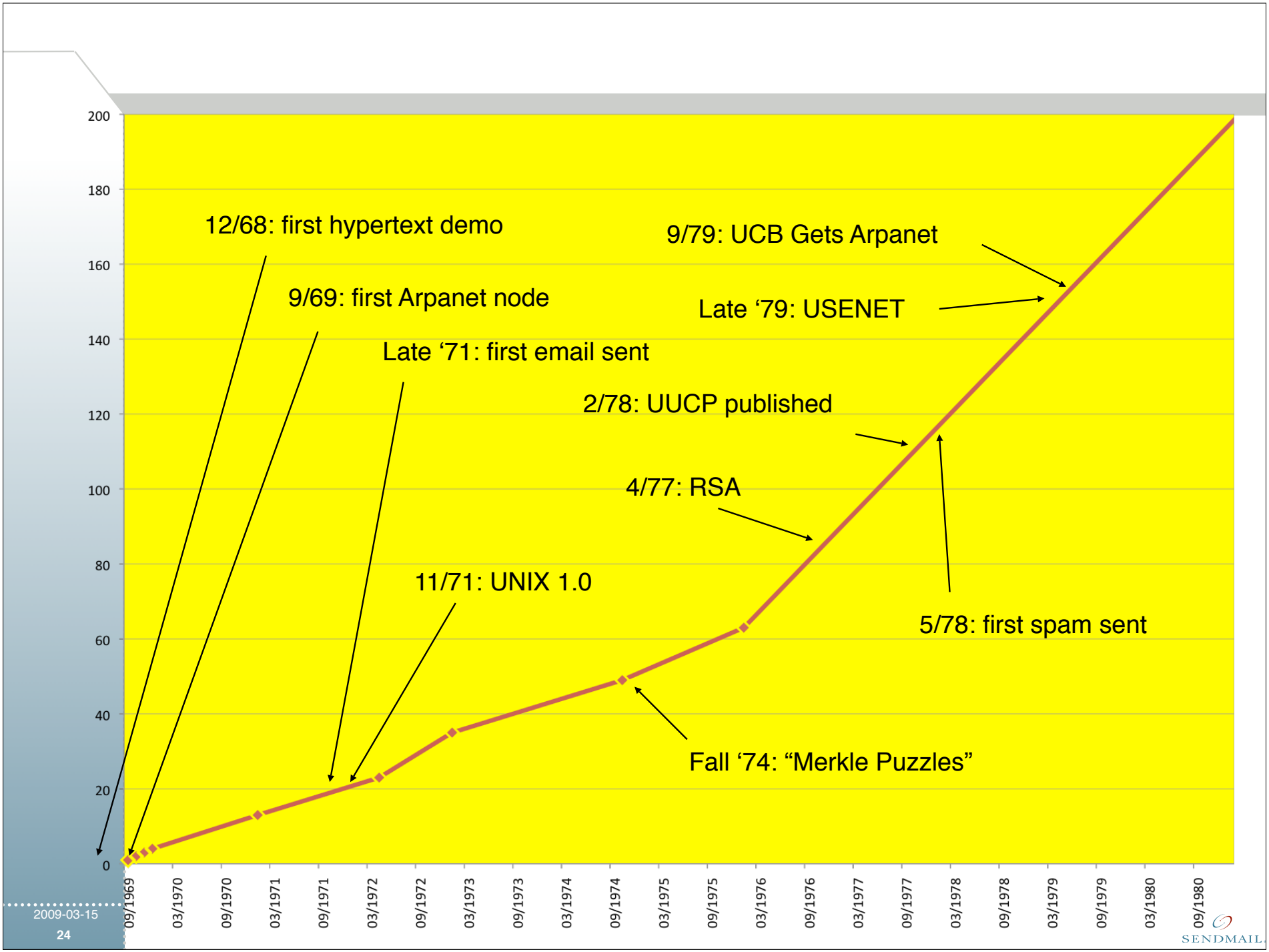
Beastie


200
180
160
140
120
100
80
60
40
20
0
12/68: first hypertext demo
9/69: first Arpanet node
Late ‘71: first email sent
11/71: UNIX 1.0
4/77: RSA
Fall ‘74: “Merkle Puzzles”
09/1969
03/1970
09/1970
03/1971
09/1971
03/1972
09/1972
03/1973
09/1973
03/1974
09/1974
03/1975
09/1975
03/1976
09/1976
03/1977
09/1977
03/1978
09/1978
03/1979
09/1979
03/1980
09/1980

SENDMAIL

Arpanet March 1977




200
180
160
140
120
100
80
60
40
20
0
12/68: first hypertext demo
9/69: first Arpanet node
Late ‘71: first email sent
11/71: UNIX 1.0
Fall ‘74: “Merkle Puzzles”
2/78: UUCP published
4/77: RSA
5/78: first spam sent
9/79: UCB Gets Arpanet
Late ‘79: USENET
09/1969
03/1970
09/1970
03/1971
09/1971
03/1972
09/1972
03/1973
09/1973
03/1974
09/1974
03/1975
09/1975
03/1976
09/1976
03/1977
09/1977
03/1978
09/1978
03/1979
09/1979
03/1980
09/1980

SENDMAIL

The ARPANET at Berkeley

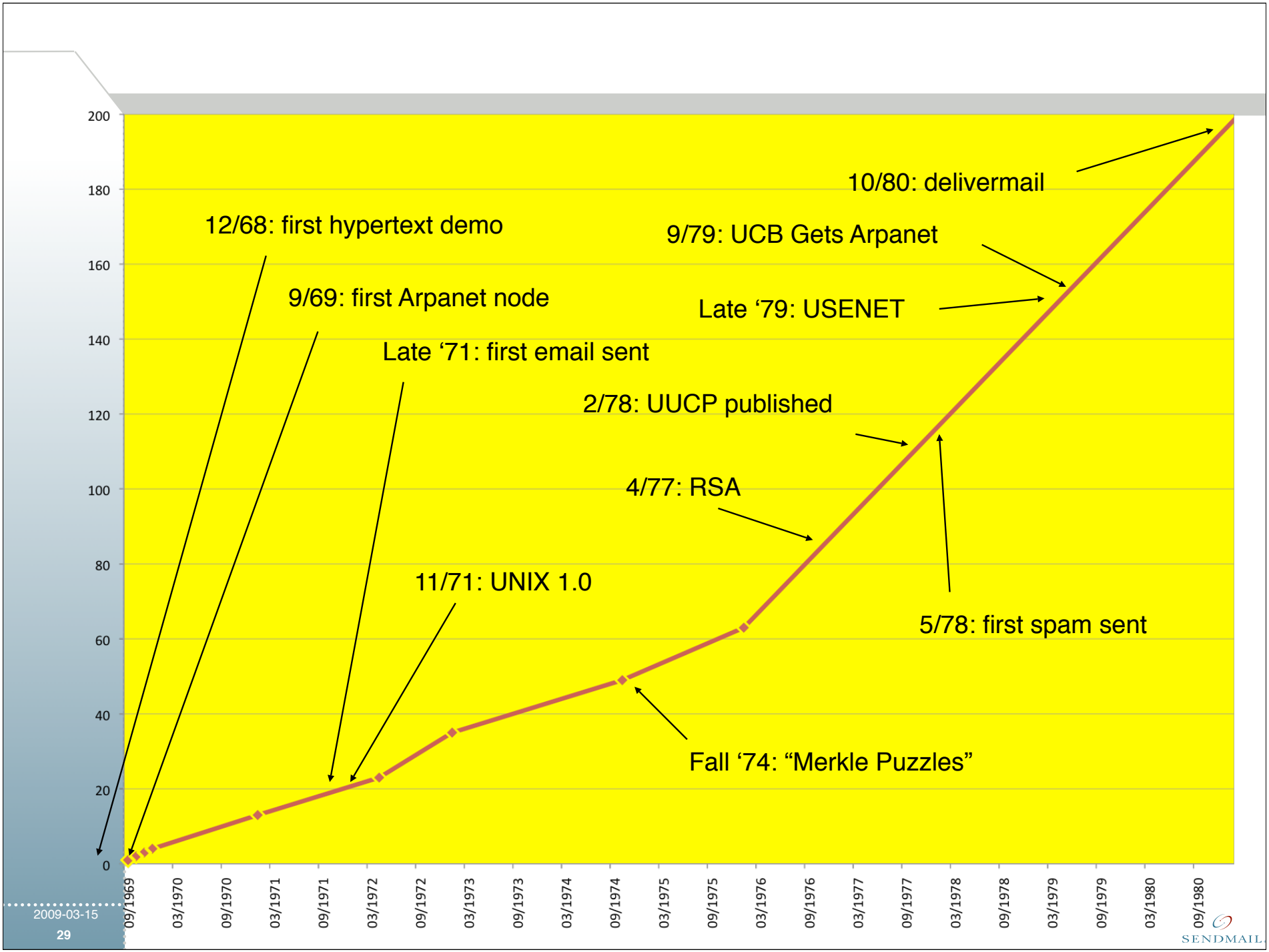
- ARPANET connection to the INGRES PDP-11/70 for Distributed Database Research (9600 baud!)
- Everyone in the CS Division wanted an account
- PDP-11 couldn't handle that many simultaneous logins (not enough memory or RS-232 ports)

A PDP-11/70 (but not ours)

When Computers had Switches and Lights…

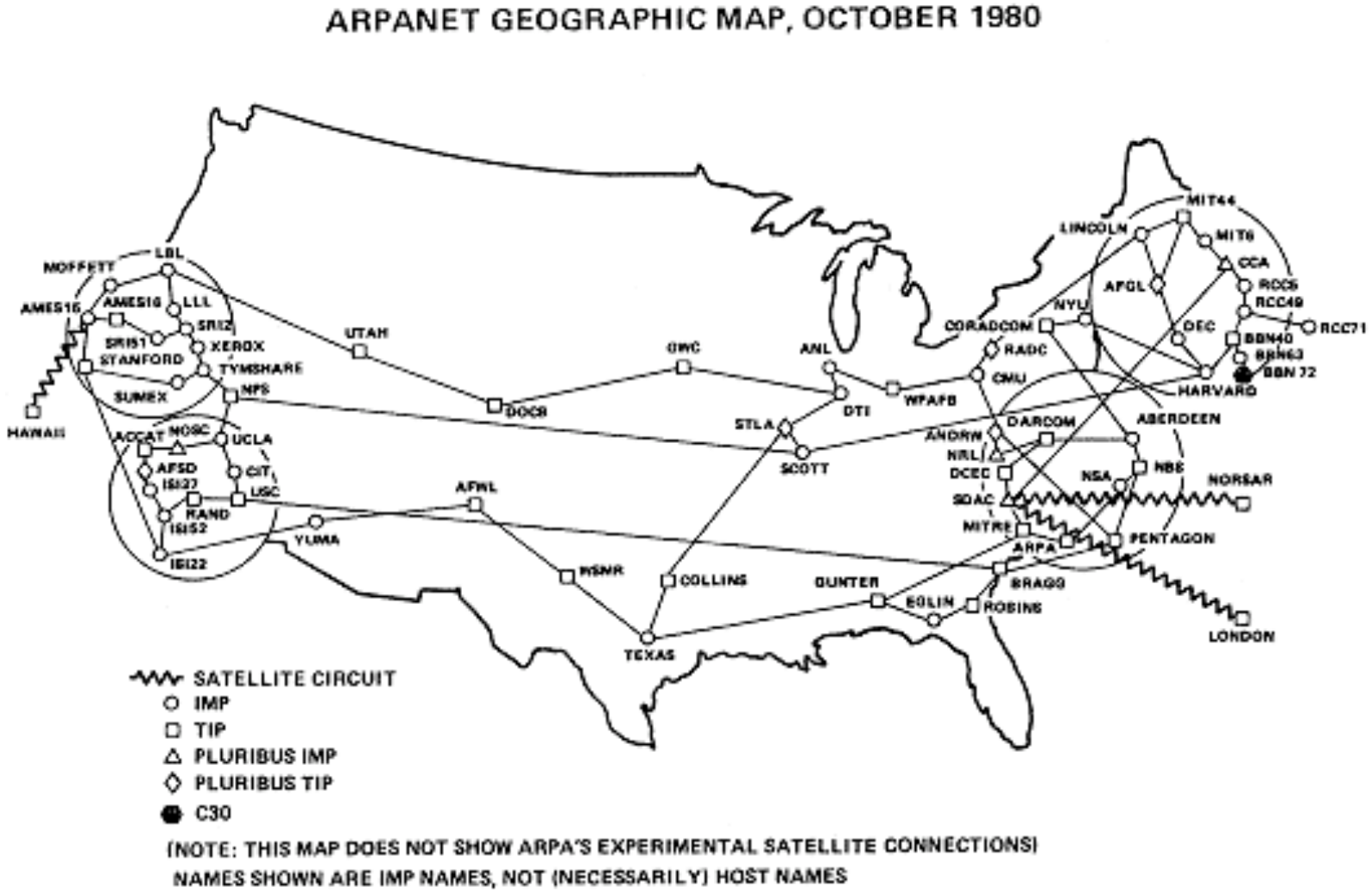
The ARPANET at Berkeley (2)

- What people really wanted was email, not full access (they didn't need telnet or FTP)
- BerkNET linked internal machines using RS-232 lines (Eric Schmidt)
- Also had a UUCP connection from Ernie CoVAX (main department machine)
- Delivermail: forward email between nets (quick hack)
- Released in 4.0 BSD (October 1980); very popular


200
180
160
140
120
100
80
60
40
20
0
09/1969
03/1970
09/1970
03/1971
09/1971
03/1972
09/1972
03/1973
09/1973
03/1974
09/1974
03/1975
09/1975
03/1976
09/1976
03/1977
09/1977
03/1978
09/1978
03/1979
09/1979
03/1980
09/1980
12/68: first hypertext demo
9/69: first Arpanet node
Late ‘71: first email sent
11/71: UNIX 1.0
Fall ‘74: “Merkle Puzzles”
4/77: RSA
2/78: UUCP published
5/78: first spam sent
9/79: UCB Gets Arpanet
Late ‘79: USENET
10/80: delivermail

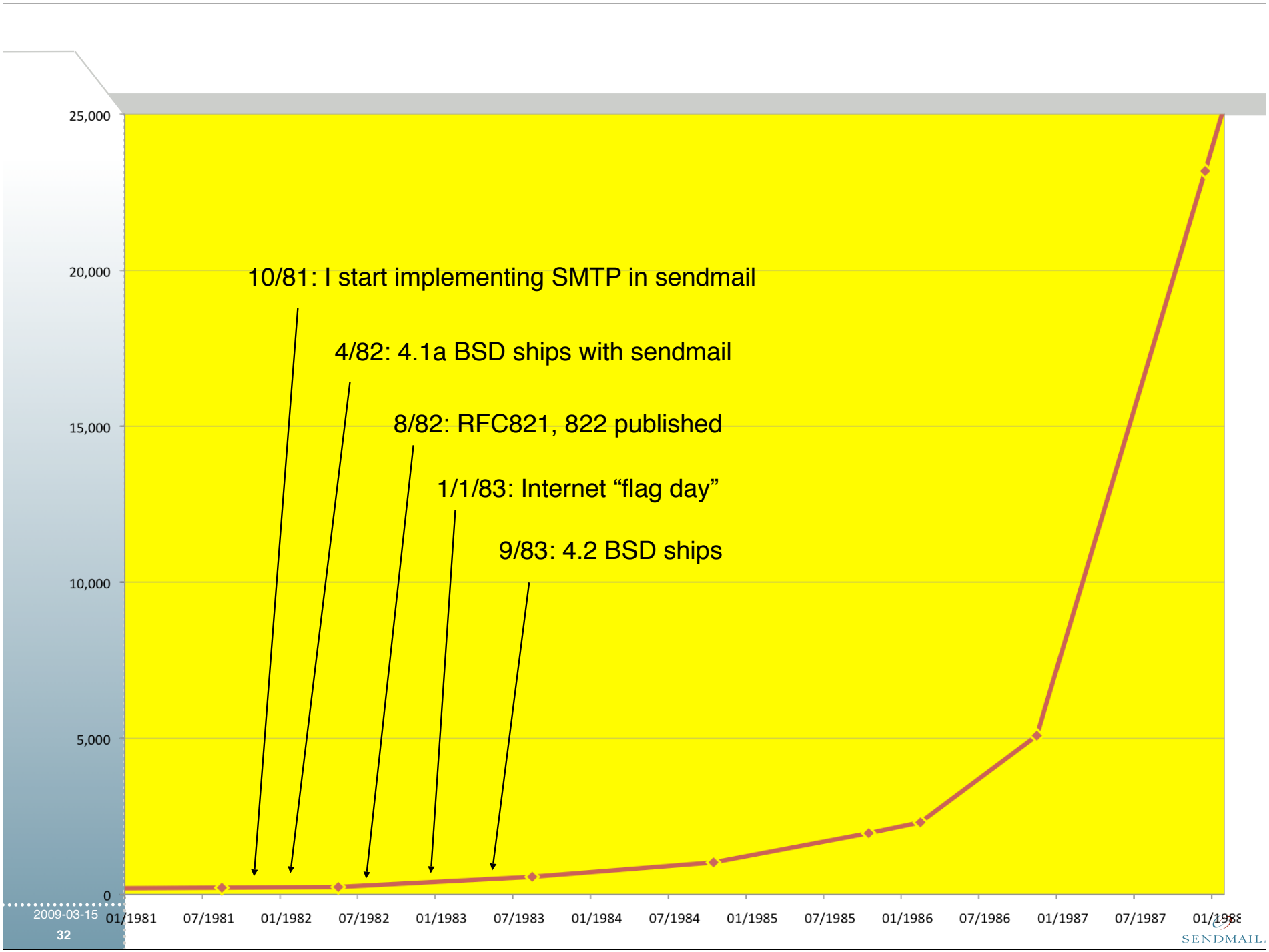
SENDMAIL

Arpanet October 1980

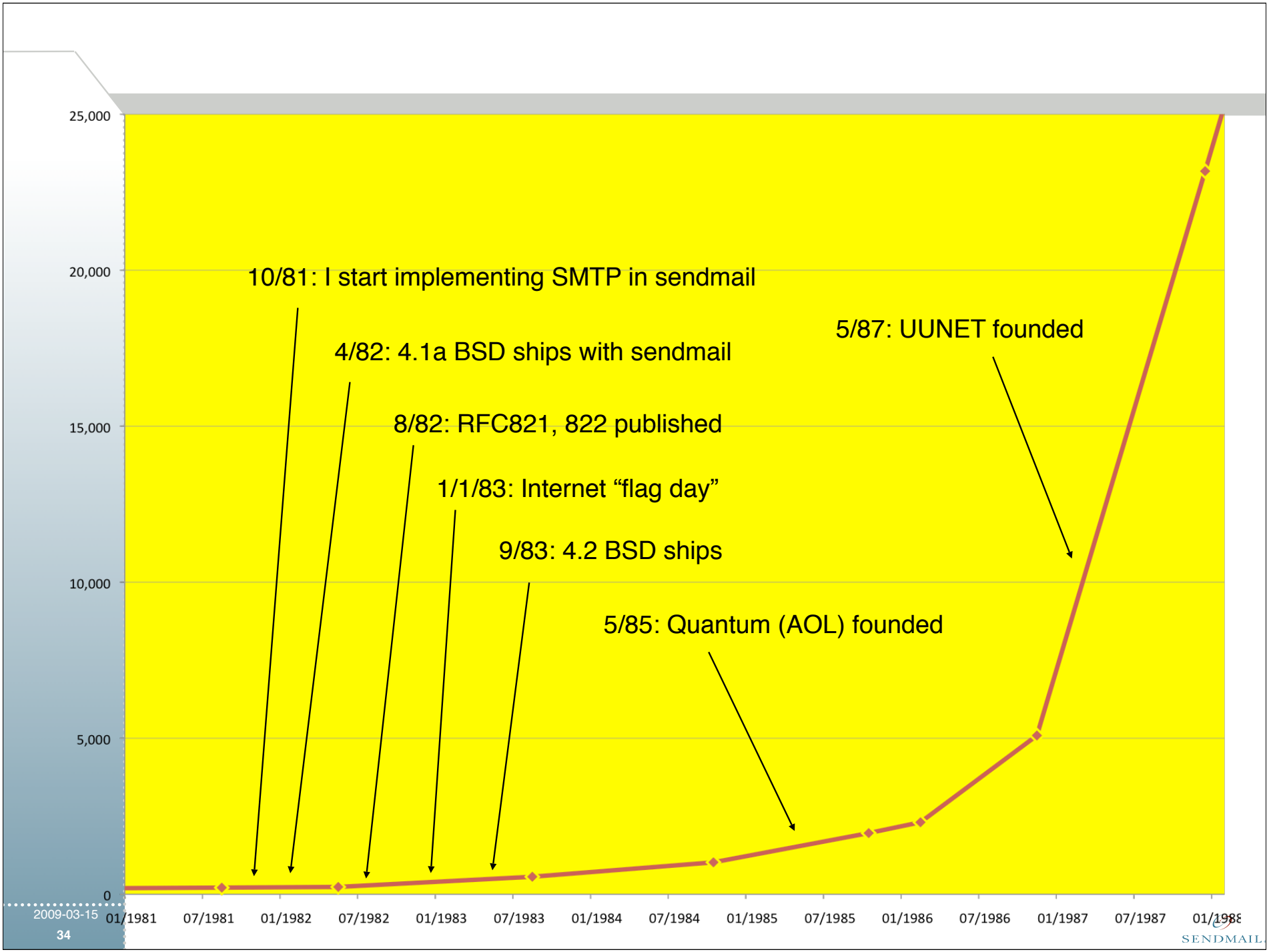


Internet History
1981–1988

213–33,000 hosts

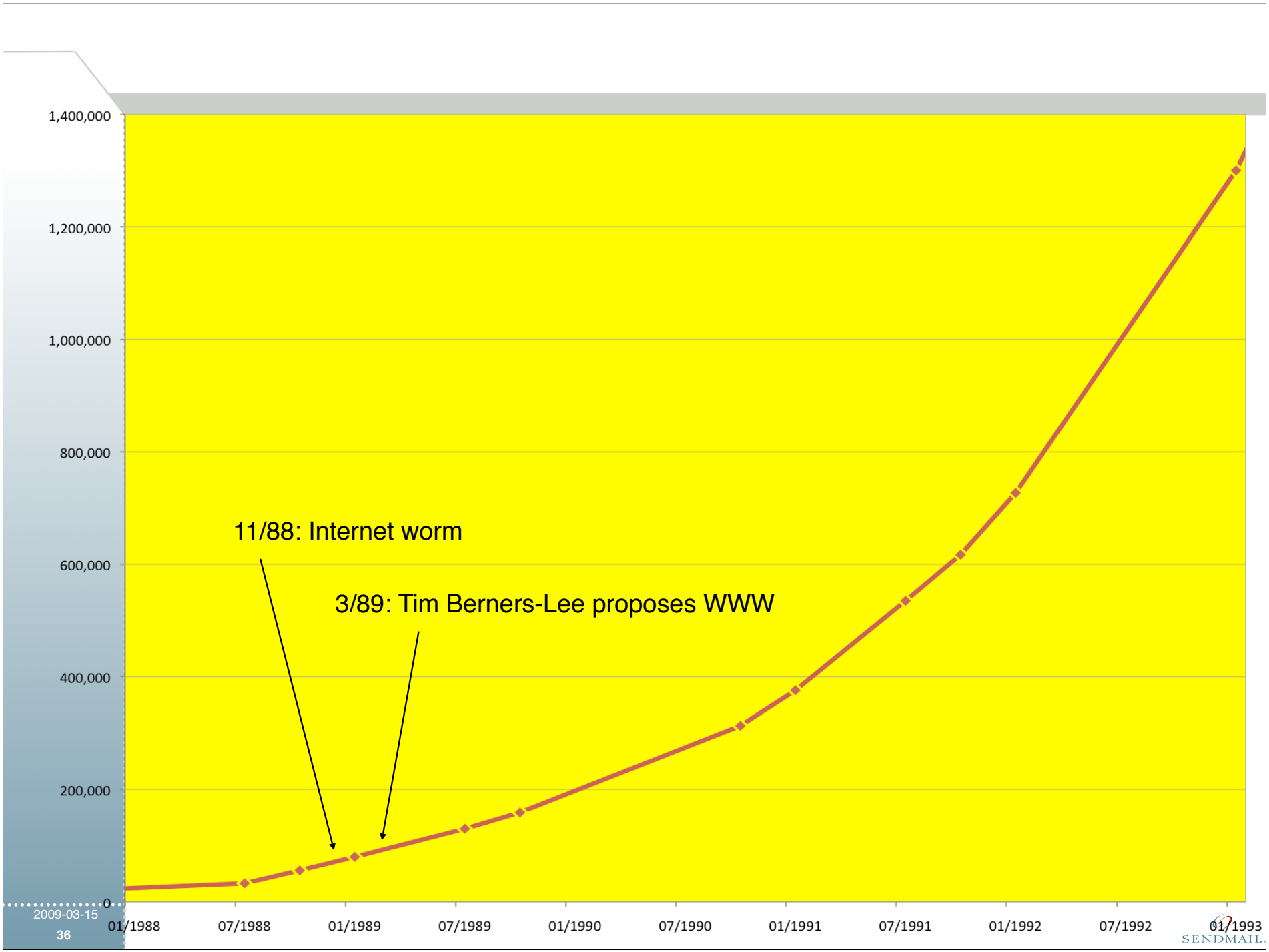

25,000
20,000
15,000
10,000
5,000
0
10/81: I start implementing SMTP in sendmail
4/82: 4.1a BSD ships with sendmail
8/82: RFC821, 822 published
1/1/83: Internet “flag day”
9/83: 4.2 BSD ships
01/1981
07/1981
01/1982
07/1982
01/1983
07/1983
01/1984
07/1984
01/1985
07/1985
01/1986
07/1986
01/1987
07/1987
SENDMAIL

Beastie in color

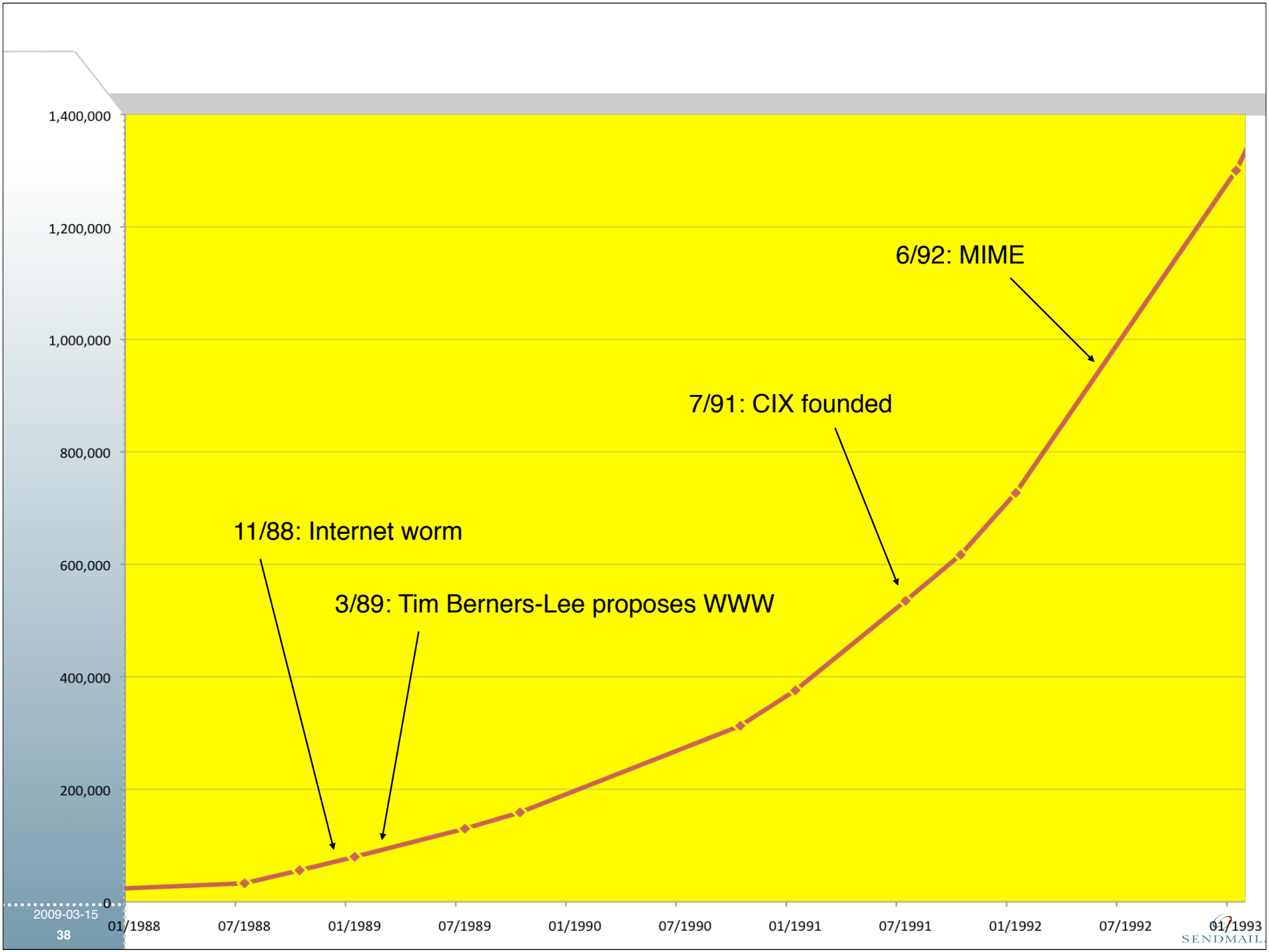

25,000
20,000
15,000
10,000
5,000
0
10/81: I start implementing SMTP in sendmail
4/82: 4.1a BSD ships with sendmail
8/82: RFC821, 822 published
1/1/83: Internet “flag day”
9/83: 4.2 BSD ships
5/85: Quantum (AOL) founded
5/87: UUNET founded
01/1981
07/1981
01/1982
07/1982
01/1983
07/1983
01/1984
07/1984
01/1985
07/1985
01/1986
07/1986
01/1987
07/1987

Internet History
1988–1993

33,000–1,300,000 hosts

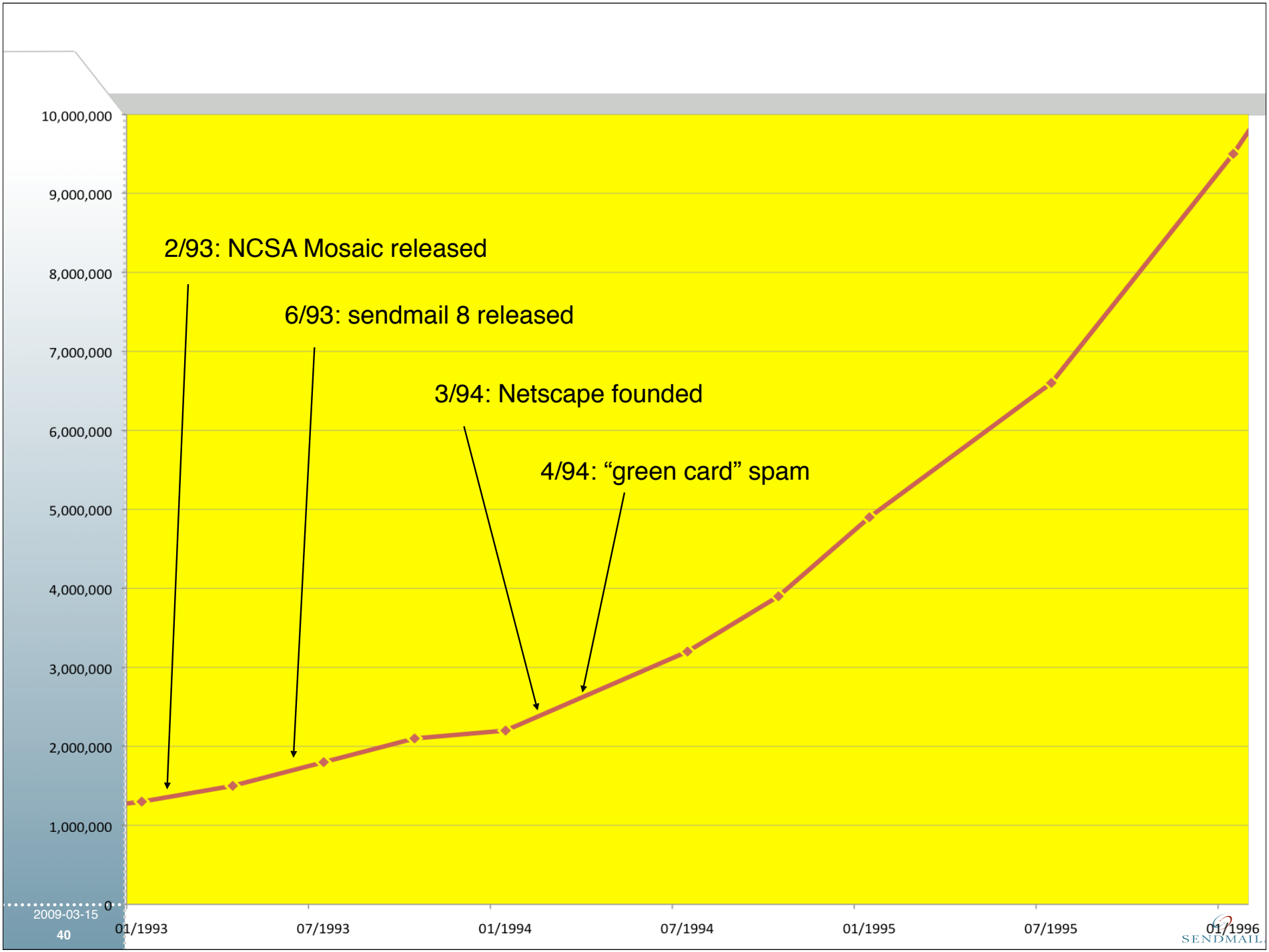

1,400,000
1,200,000
1,000,000
800,000
600,000
400,000
200,000
0
11/88: Internet worm
3/89: Tim Berners-Lee proposes WWW
01/1988
07/1988
01/1989
07/1989
01/1990
07/1990
01/1991
07/1991
01/1992
07/1992
01/1993

Tim Berners-Lee


1,400,000
1,200,000
1,000,000
800,000
600,000
400,000
200,000
0
6/92: MIME
7/91: CIX founded
11/88: Internet worm
3/89: Tim Berners-Lee proposes WWW
01/1988
07/1988
01/1989
07/1989
01/1990
07/1990
01/1991
07/1991
01/1992
07/1992
01/1993

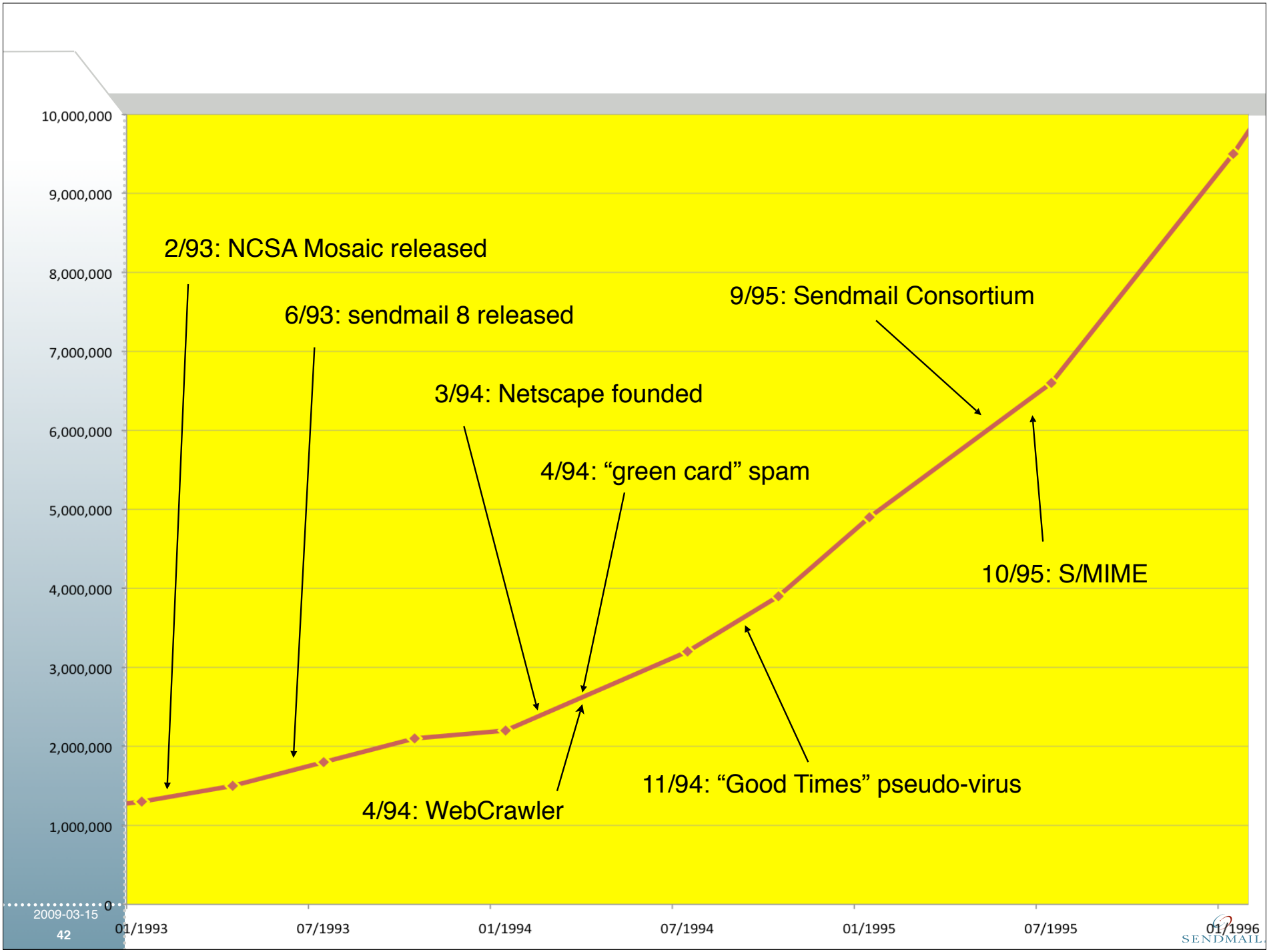
Internet History
1993–1996

1,300,000–12,900,000 hosts


10,000,000
9,000,000
8,000,000
7,000,000
6,000,000
5,000,000
4,000,000
3,000,000
2,000,000
1,000,000
0
2/93: NCSA Mosaic released
6/93: sendmail 8 released
3/94: Netscape founded
4/94: “green card” spam
01/1993
07/1993
01/1994
07/1994
01/1995
07/1995
01/1996

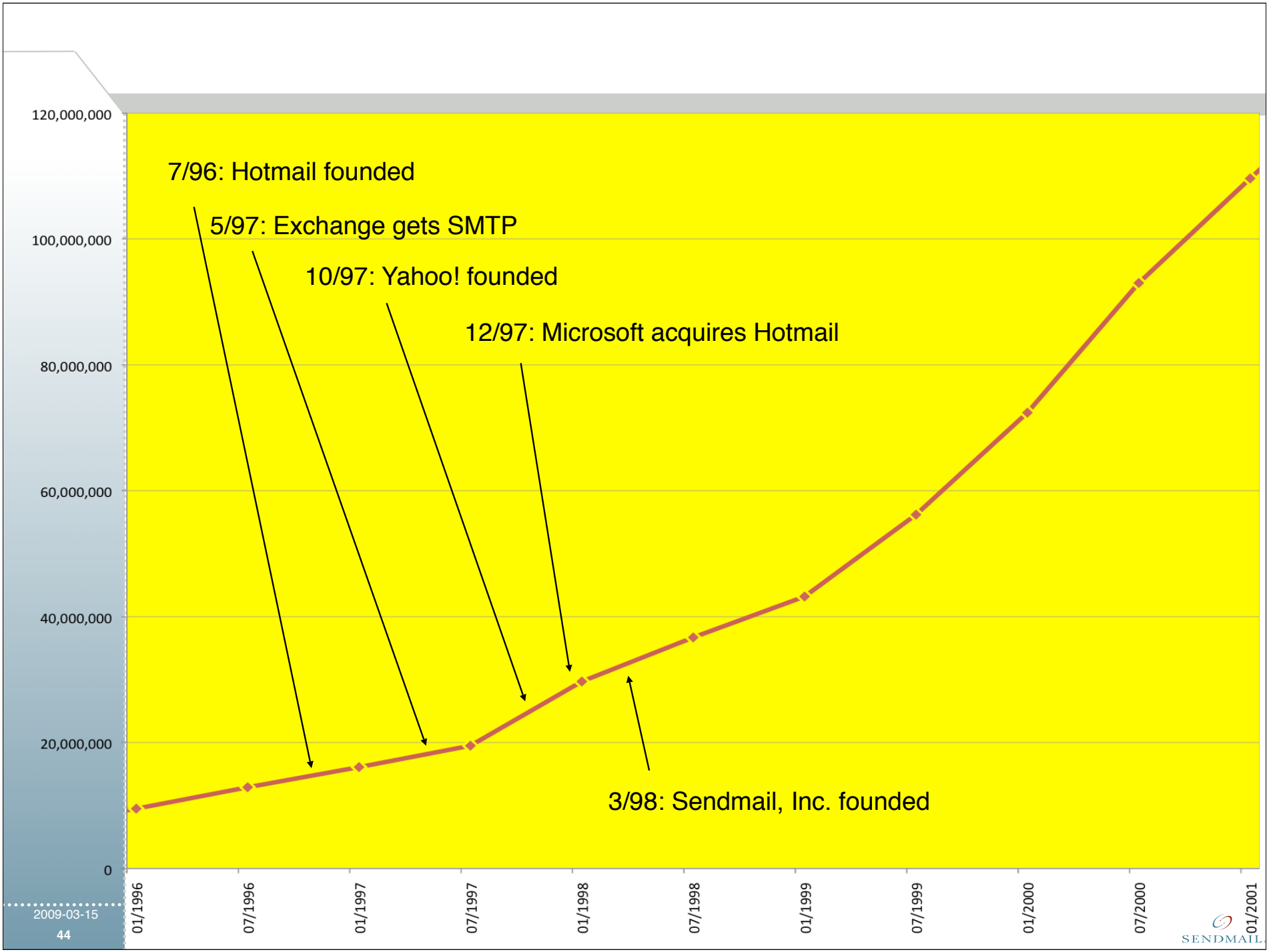
Spam

- Network previously largely cooperative
- Flame wars all too common but isolated
- Spam not new: September 13, 1904 — unsolicited commercial email via telegraph
- DEC spam: May 1, 1978
- “Make Money Fast” chain letter: 1988
- Attitudes change: “Greed is Good” comes to the Internet
- Canter and Siegel (Green Card Spam) were unapologetic

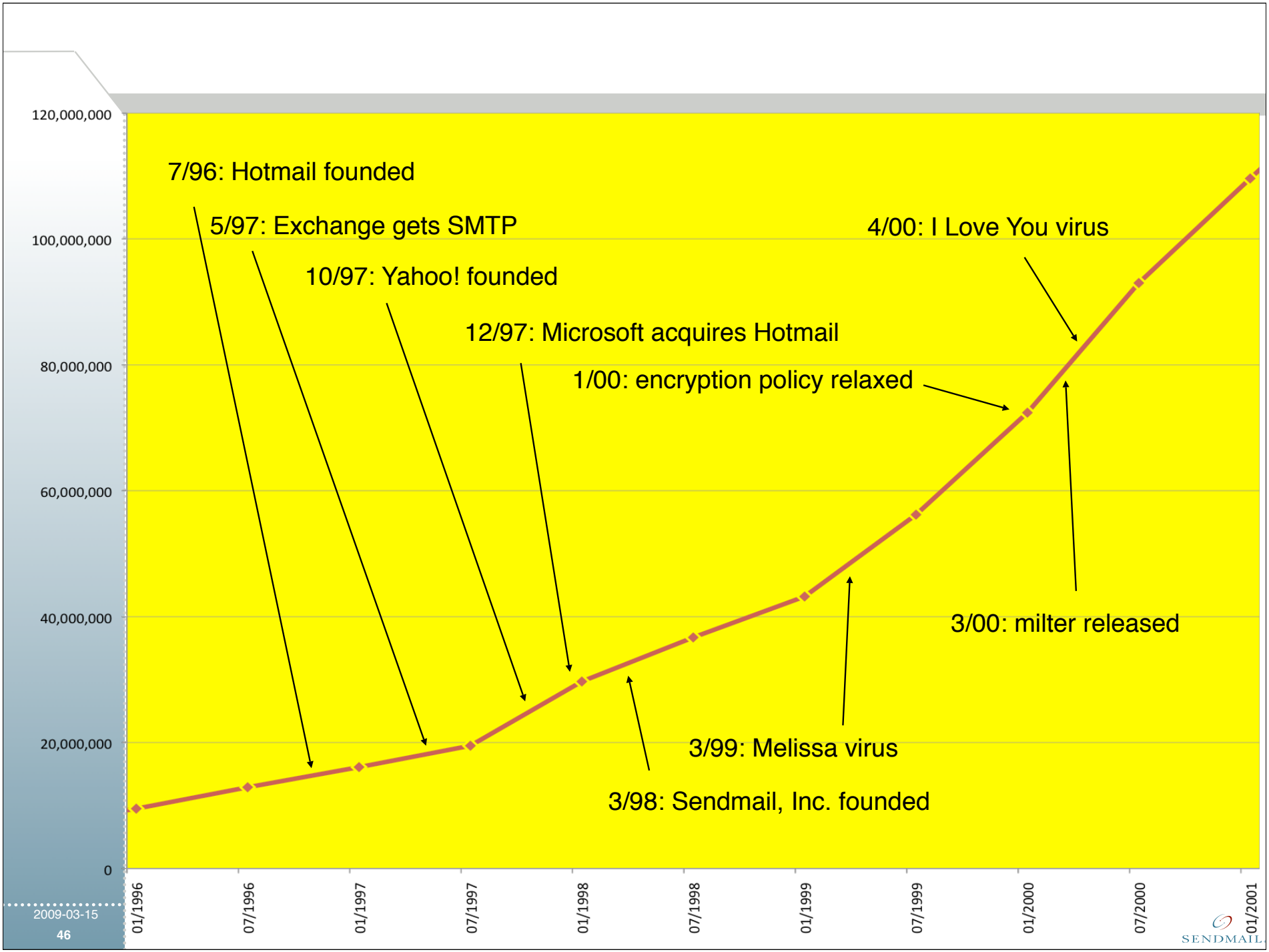

10,000,000
9,000,000
8,000,000
7,000,000
6,000,000
5,000,000
4,000,000
3,000,000
2,000,000
1,000,000
0
2/93: NCSA Mosaic released
6/93: sendmail 8 released
3/94: Netscape founded
4/94: "green card" spam
4/94: WebCrawler
11/94: "Good Times" pseudo-virus
9/95: Sendmail Consortium
10/95: S/MIME
01/1993
07/1993
01/1994
07/1994
01/1995
07/1995
01/1996

Internet History
1996–2000

12,900,000–109,000,000 hosts

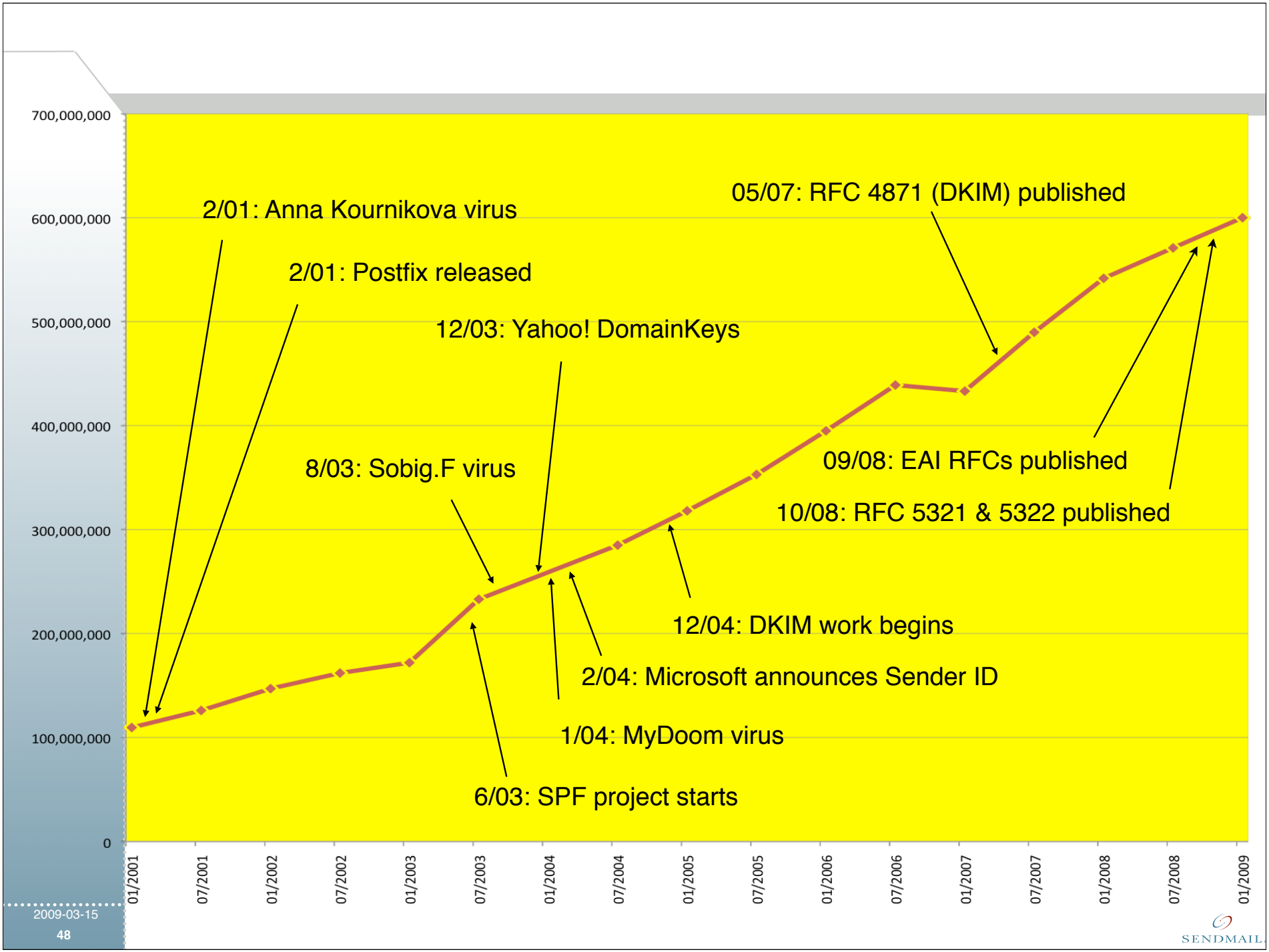

120,000,000
100,000,000
80,000,000
60,000,000
40,000,000
20,000,000
0
7/96: Hotmail founded
5/97: Exchange gets SMTP
10/97: Yahoo! founded
12/97: Microsoft acquires Hotmail
3/98: Sendmail, Inc. founded
01/1996
07/1996
01/1997
07/1997
01/1998
07/1998
01/1999
07/1999
01/2000
07/2000
01/2001
SENDMAIL

Sendmail, Inc.


120,000,000
100,000,000
80,000,000
60,000,000
40,000,000
20,000,000
0
7/96: Hotmail founded
5/97: Exchange gets SMTP
10/97: Yahoo! founded
12/97: Microsoft acquires Hotmail
1/00: encryption policy relaxed
4/00: I Love You virus
3/00: milter released
3/99: Melissa virus
3/98: Sendmail, Inc. founded
01/1996
07/1996
01/1997
07/1997
01/1998
07/1998
01/1999
07/1999
01/2000
07/2000
01/2001
SENDMAIL

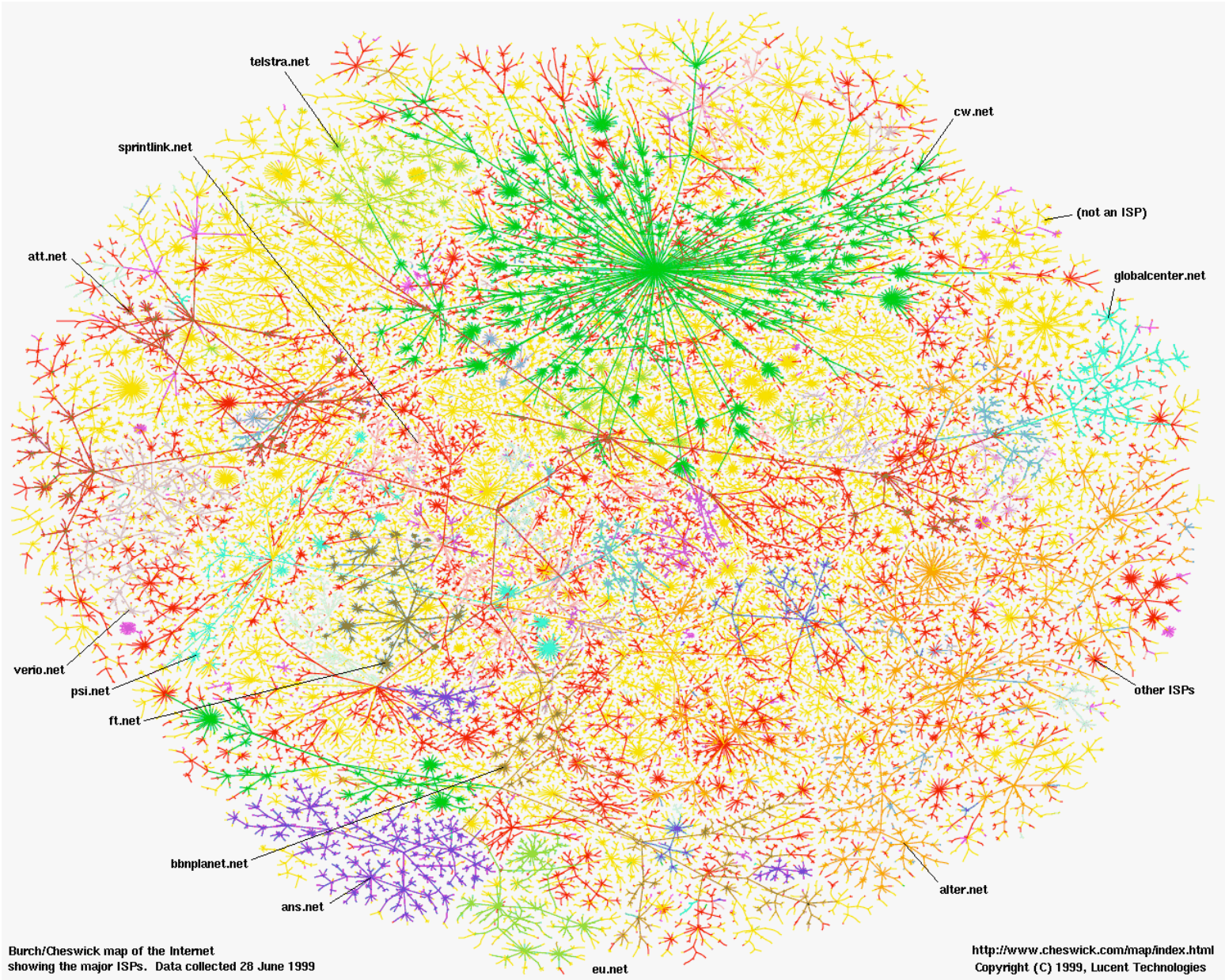
Internet History
2001–2009

109,600,000–439,000,000 hosts


700,000,000
600,000,000
500,000,000
400,000,000
300,000,000
200,000,000
100,000,000
0
2/01: Anna Kournikova virus
2/01: Postfix released
12/03: Yahoo! DomainKeys
8/03: Sobig.F virus
05/07: RFC 4871 (DKIM) published
09/08: EAI RFCs published
10/08: RFC 5321 & 5322 published
12/04: DKIM work begins
2/04: Microsoft announces Sender ID
1/04: MyDoom virus
6/03: SPF project starts
01/2001
07/2001
01/2002
07/2002
01/2003
07/2003
01/2004
07/2004
01/2005
07/2005
01/2006
07/2006
01/2007
07/2007
01/2008
07/2008
01/2009

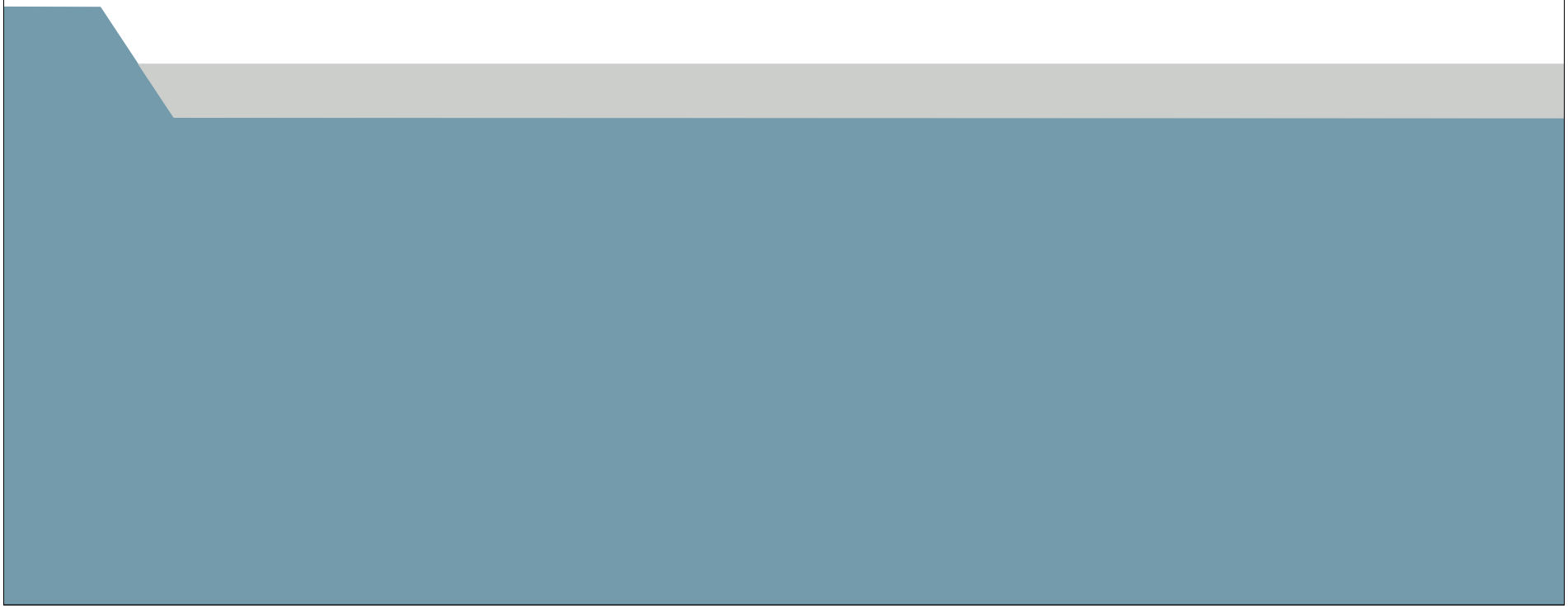
SENDMAIL

Internet June 1999



ENDMAIL

Messaging, Spam, Security, and Authentication



Email Spamming

- Economic issue: C(sending) ≪ C(receiving)
  - Note: cost function C is not denominated in $ or ¥
  - Can be CPU time, memory use, human time, etc.
- Possible approaches
  - Content filtering (reaching limits; doesn't fix cost function)
  - ePostage (infrastructure & acceptance problems)
  - Challenge-Response (poor scaling; user confusion)
  - HashCash (useless with zombie farms)
  - Graylisting (easy to defeat; broken by server farms)
  - Authentication (insufficient by itself)

Phishing

- Try to trick someone into giving up private information, sending money, etc.
- Not limited to email
- Nothing new, just easier to do
  - “Nigerian” or “419” attacks predated email
  - Shysters prey on elderly
- “Shotgun” phishing overlaps with spam
- “Spear phishing” changes the rules
  - Can spend significant money targeting someone
  - They usually have significant information about you
- Attackers often try to pretend to be someone you trust
  - Authentication can really help with this

Issues with Authentication

- Choosing the standard(s)
  - Path-based (SenderID/SPF)
  - Signature-based (DKIM, DomainKeys)
  - Neither authenticates users, just SMTP nodes
  - Both can break on non-malicious messages
- How to achieve adoption?
  - Sender- or Recipient-driven?
  - Senders seem to be the primary driver
- Status of unsigned email
  - Unsigned mail must remain legal during transition
  - Author Domain Signing Practices (ADSP) tells how to treat unsigned mail (DKIM only)
- Authentication by itself is not enough

DKIM details

- DKIM contains two specifications
  - “Base” signing specification — how to sign an individual message
  - “Signing Practices” — how to interpret unsigned messages
- Base spec (RFC4871) published May 2007
  - Signs body and selected headers using keypairs
  - Public key management done in DNS
- Signing Practices is very controversial
  - Even the name changed a few times — ultimate name: Author Domain Signing Practices (ADSP)
  - Finally got watered down to the point where there was nothing controversial left

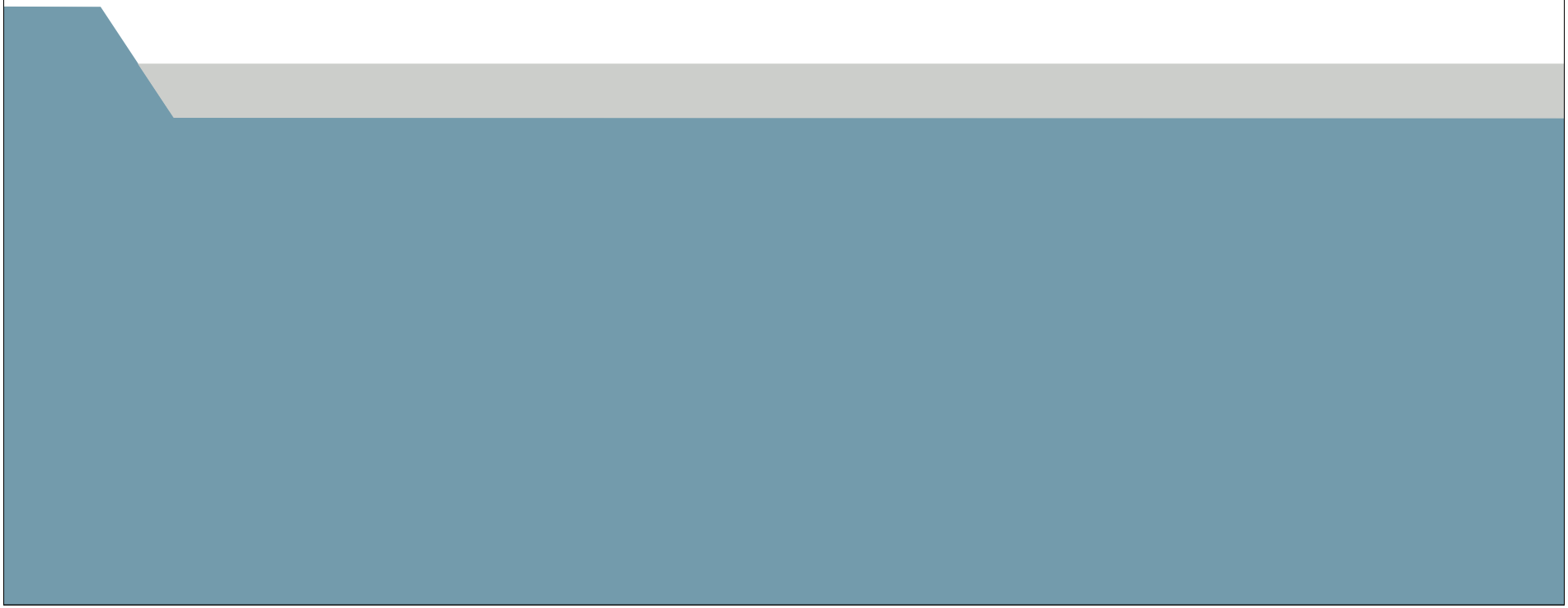
Issues with Reputation

- Negative reputation is well known
  - E.g., DNS blackhole lists
  - Can change very quickly
- Positive reputation is hard without "breaking" the world we know today
  - Does "presumption of innocence" become "presumption of guilt"?
  - How does the smaller player join the club?
- Will we go to a "closed society" email model?
  - New domains will have no reputation, so recipients may be unwilling to accept their mail
  - This is (sort of) what Challenge-Response does
- Accreditation for a fee?

Other Messaging Not Immune from Abuse

- SPIM (Instant Messaging Spam) becoming more common
  - Authentication helps, but too easy to get accounts
- Social Networking sites being targeted
  - MySpace and Facebook have both been targeted
  - Often successful because messages seem to come from “friends”
  - Particularly good for phishing
- SPIT (Internet Telephony Spam) is a growth industry

Some Speculations on the Future



Email of the Future — Short Term Predictions

- Authenticated email with (some) sender reputation, growing with time
- Slow movement toward Email Address Internationalization (EAI IETF Working Group)
  - Downgrading is very hard to do and may not happen at all
- Young people moving off SMTP-based email, using IM, Social Networking, micro-blogging instead

Email of the Future — Medium Term Predictions

- SMTP is dead, long live SMTP (?)
  - Often proposed and predicted, but so far SMTP survives
  - Could fix some problems, but…
  - SMTP (with extensions) is “good enough”
- UTF-8 everywhere
  - Probably will happen; not hard to do (except for downgrading)
- Most email will be at hosted providers
  - Getting ever harder to build and operate a mail system due to increasing challenges and demands

Random Longer-Term Predictions

- Casual messaging will continue to move away from traditional email
  - People crave instantaneous gratification
  - Fits better with mobile usage
- Email will not die
  - Too well suited to business
  - Need for longer, more considered messages
  - Security and regulatory constraints
  - Cheaper than texting (for now; this cost is artificial)
  - IM doesn't work well across time zones
- Distinction between email, IM, voice will blur
- Interesting work: Computer/Human interface

The Unified Messaging Mess (non-email)

- Too many messaging services and applications with too much overlap, not enough compat

Service /App	IM	SMS	Voice	Video	Inc Calls	Out Calls	Call Fwd	VM	Addr Book	Origin
Adium	✔	✔	✘	✘	✘	✘	✘	✘	✔	IM
Skype	✔	✚	✔	✔	✚	✚	✔	✔	✔	Voice
Google Voice	✘③	✔	✔	✘	✔	US ✔ Int ✚	✔ ①	✔	✔	Call Forward
Broad voice	✘	✘	✔	✘	✔②	✔	✔	✔	✔	VoIP, $
Gizmo	✔	✔	✔	✔	✚	✚	✔	✔	✔	VoIP soft
iChat	✔	✔	✔	✔	✘	✘	✘	✘	✔	IM

ernie:ucbvax!decwrl!x2300::mike@berkeley.edu

Thank You

Gratuitous Advertising:

Conference on Email and Anti-Spam (CEAS)

July 16–17, 2009

Mountain View, California

http://ceas.cc