Confessions of an Internet Timekeeper

David L. Mills
University of Delaware
http://www.eecis.udel.edu/~mills
mailto:mills@udel.edu

alautun, Maya glyph

On the Internet cultural evolution

- “We have met the enemy and he is us.” – Walt Kelly.
- Maybe the most important lesson of the Internet was that the technology was developed and refined by its own users.
- There was a certain ham-radio mentality where users/developers had great fun making new protocols to work previously unheard applications.
- The developers were scattered all over the place, but they had a big, expensive sandbox with little parental supervision.
- There is no doubt that the enthusiasm driving the developers was due to the urgent need to communicate with each other without wasting trees or airplane fuel.

Reality check

- The primary motivation for the Internet model was the need for utmost reliability in the face of untried hardware, buggy programs and lunch.
  - The most likely way to lose a packet is a program bug, rather than a transmission error.
  - Something somewhere was/is/will always be broken at every moment.
  - The most trusted state is in the endpoints, not inside the network.
- Meanwhile, we need to timestamp things.
  - Mostly, this was for distributed experiment and performance evaluation.
  - It became even more useful as Internet seismograph.

Introduction

- We talk here about synchronizing computer clocks in the Internet and its tributaries.
- An estimated 25 million Network Time Protocol (NTP) servers and clients are deployed all over the world.
- NTP software has been ported to almost every workstation and server platform available today - from PCs to supercomputers and embedded systems, even home routers and battery backup systems.
- In the remainder of this talk we discusses the technology and lessons learned in its deployment and use.

The Sun never sets on NTP

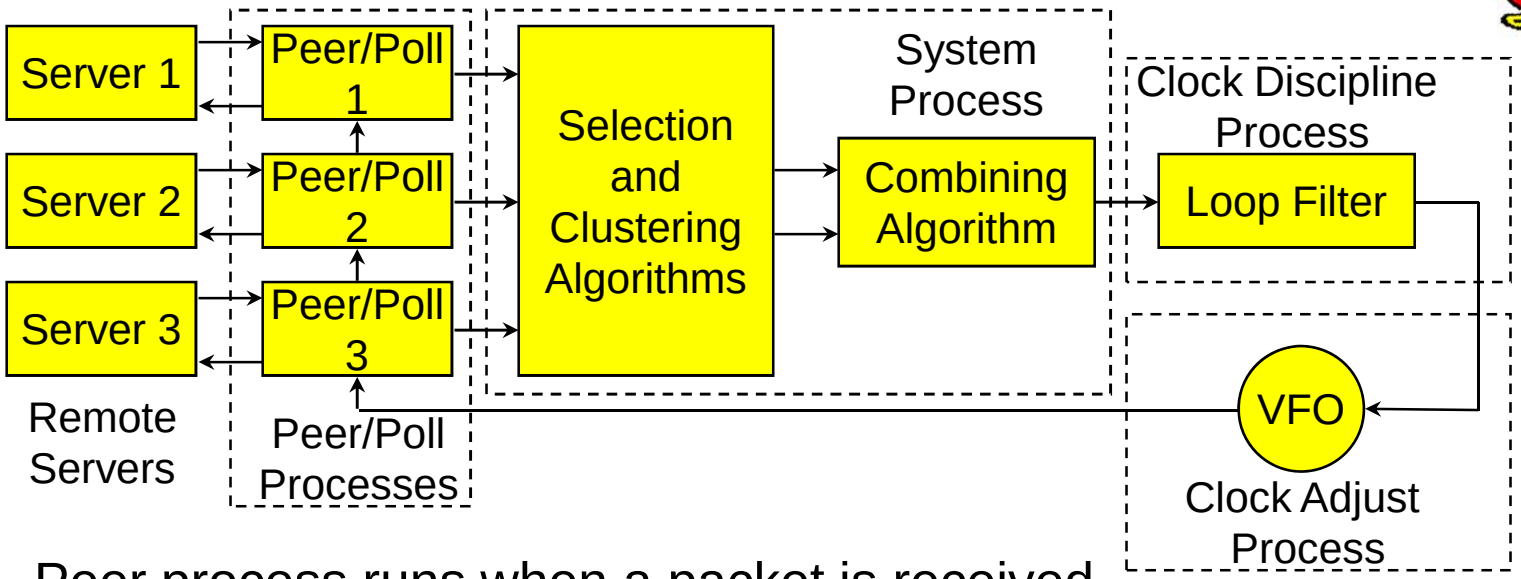
- NTP is arguably the longest running, continuously operating, ubiquitously available protocol in the Internet (since 1979)
  - In the US, USNO and NIST operate multiple public NTP primary servers directly synchronized to national standard cesium clock ensembles and GPS.
  - Government agencies in many other countries and on all continents (including Antarctica) operate public NTP primary servers.
  - National and regional service providers operate public NTP secondary servers synchronized to the primary servers.
  - US Government agencies, including US Weather Service, US Treasury Service, IRS and FAA operate their own NTP networks.
  - Private and public institutions, including universities, broadcasters, financial institutions and corporations operate their own NTP networks.
  - NTP is on the seabed, Navy warships, NASA Shuttle missions and planned for the Mars Internet.

Background and disclaimer

- This is a personal retrospective, not a history archive, and covers topics important to me and which were my major research interests.
  - From the perspective of the Internet program managers, I was the internet greasemonkey. I built things and tested them, often with the help of NTP.
  - I chaired the Gateway Algorithms and Data Structures (GADS) and later the Internet Architecture (INARC) task forces and was a member of the Internet Control and Configuration Board (ICCB) and later the Internet Activities Board (IAB).
  - On my watch was gateway architecture, network and internetwork routing algorithms, timekeeping and growing pains.
  - Along the way I developed an obsession to wind every computer clock on the planet.

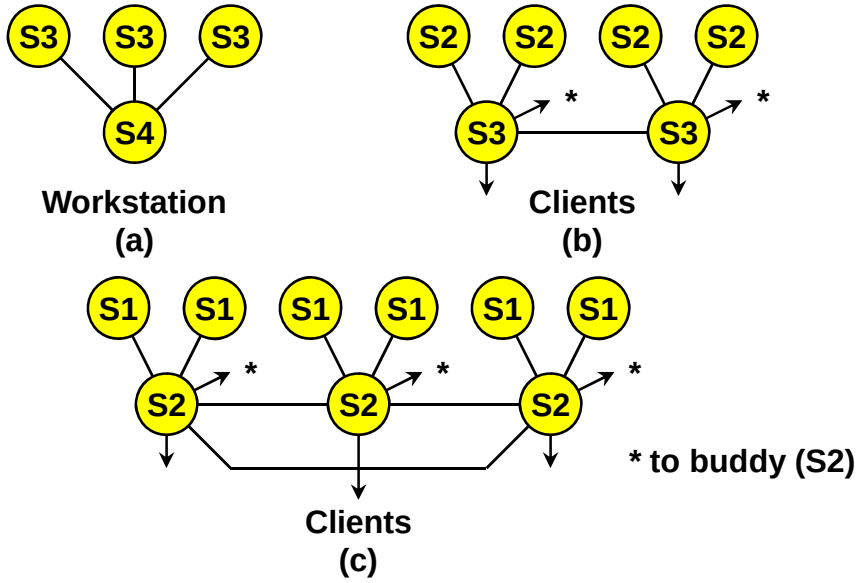
Stuffed gator presented to this alligator hunter for his help in draining the swamp.

Architectural overview



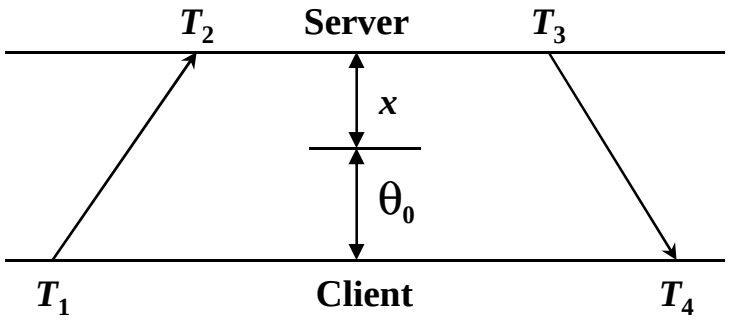
- Peer process runs when a packet is received.
- Poll process sends packets at intervals determined by the clock discipline process and remote server.
- System process runs when a new update is received.
- Clock discipline process implements clock time adjustments.
- Clock adjust process implements periodic clock frequency (VFO) adjustments.

NTP clients and servers

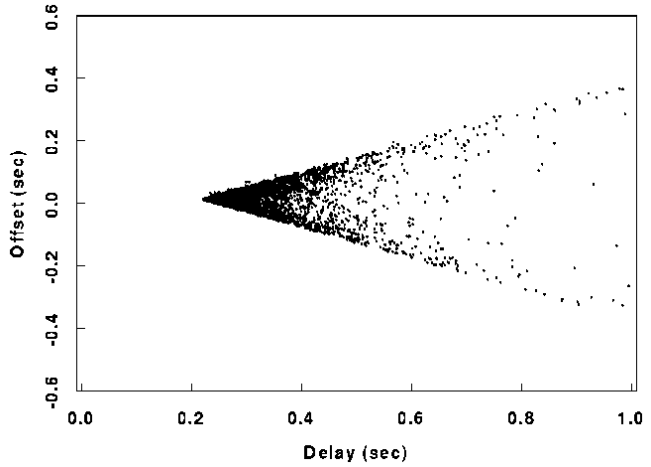


- (a) Workstations use multicast mode with multiple department servers
- (b) Department servers use client/server modes with multiple campus servers and symmetric modes with each other
- (c) Campus servers use client/server modes with up to six different external primary servers and symmetric modes with each other and external secondary (buddy) servers

Clock filter algorithm

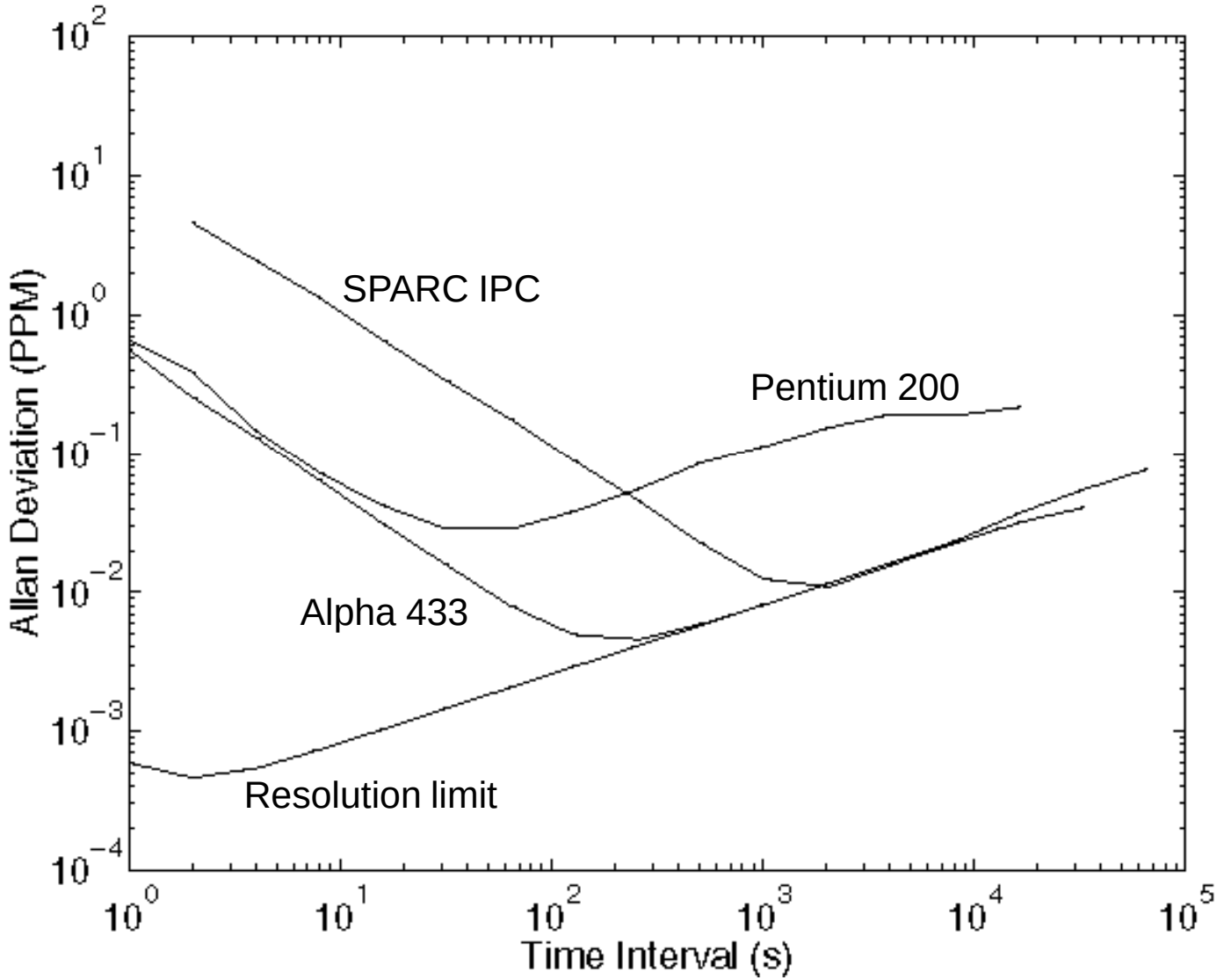


$$\theta = \tfrac{1}{2}[(T_2 - T_1) + (T_3 - T_4)]$$
$$\delta = (T_4 - T_1) - (T_3 - T_2)$$



- The most accurate offset θ_0 is measured at the lowest delay δ_0 (apex of the wedge scattergram).
- The correct time θ must lie within the wedge $\theta_0 \pm (\delta - \delta_0)/2$.
- The δ_0 is estimated as the minimum of the last eight delay measurements and (θ_0, δ_0) becomes the peer update.

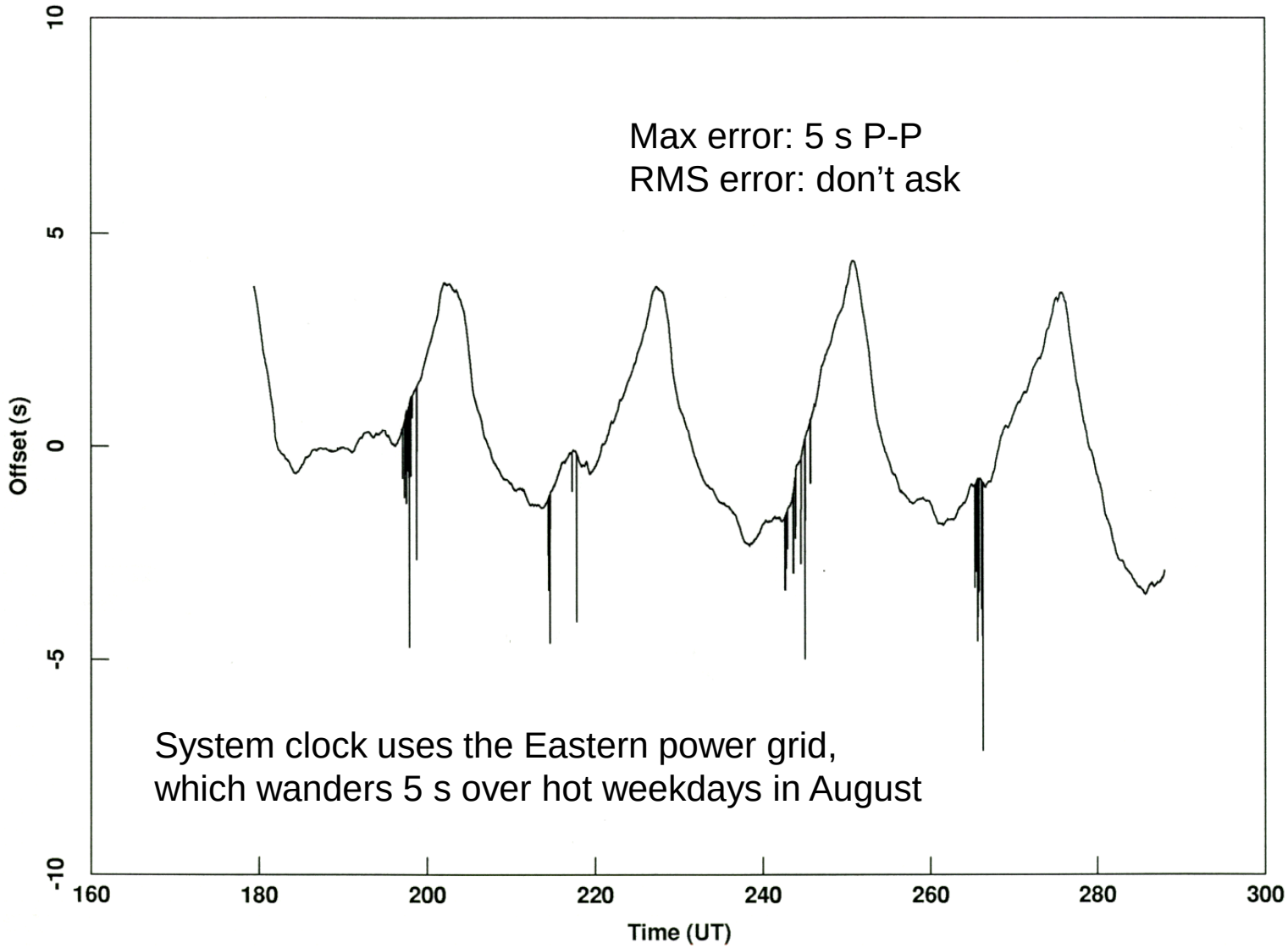
Allan deviation for typical computer clocks



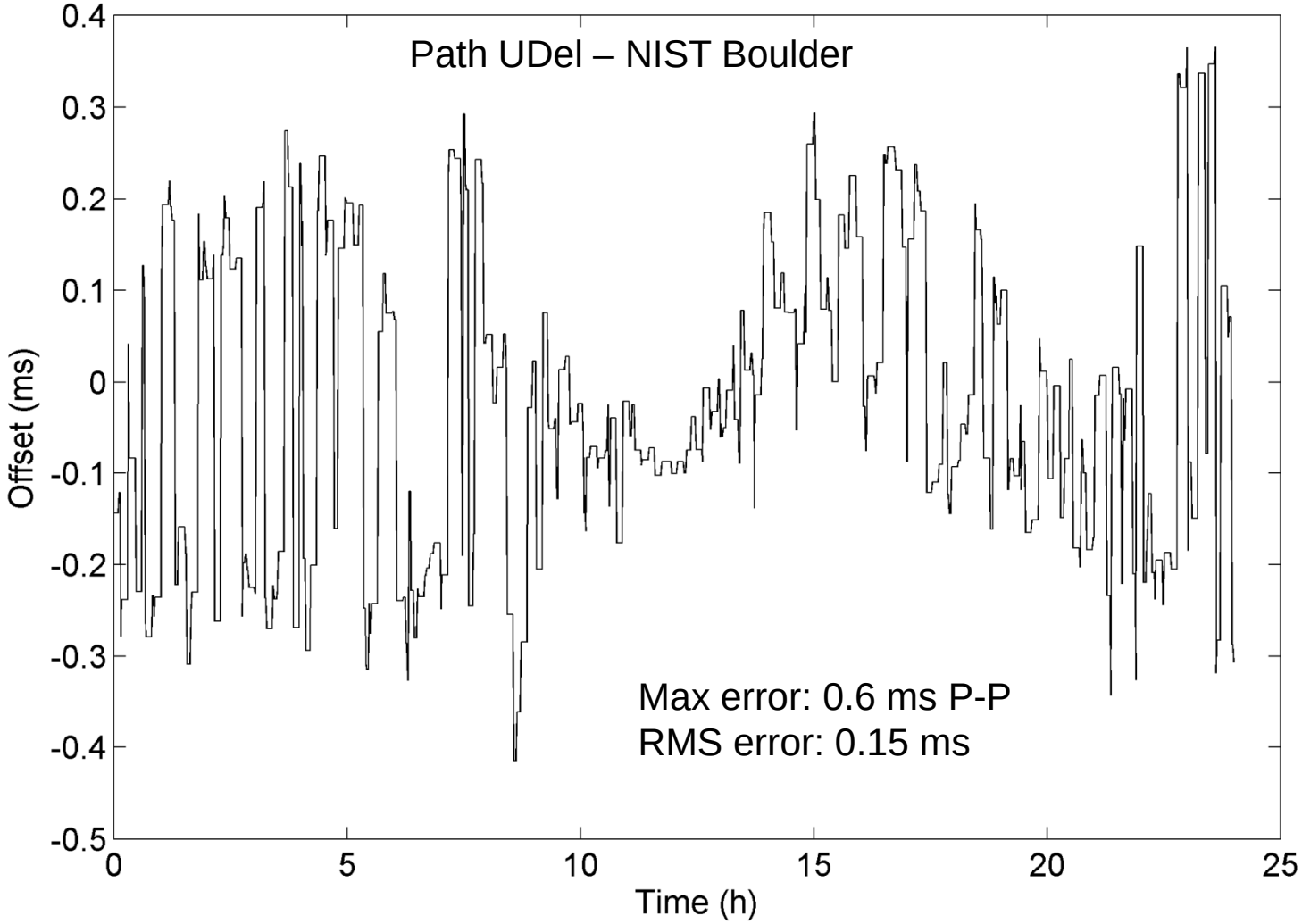
Computer clock modelling

- The vee-shaped curves on the previous slide show the Alan deviation which characterizes each individual computer clock oscillator.
- The curves show the frequency differences averaged over increasing intervals in log-log scales.
- The downward-tending traces with slope –1 are due to white phase noise as the result of network jitter.
- The upward-tending traces with slope +0.5 are due to random-walk frequency noise as the result of oscillator frequency wander.
- The intersection of the downward and upward traces is called the Allan intercept. It represents the optimum averaging interval, which results in a poll interval in the range 40-100 s.
- NTP normally operates on the high side to minimize network load.

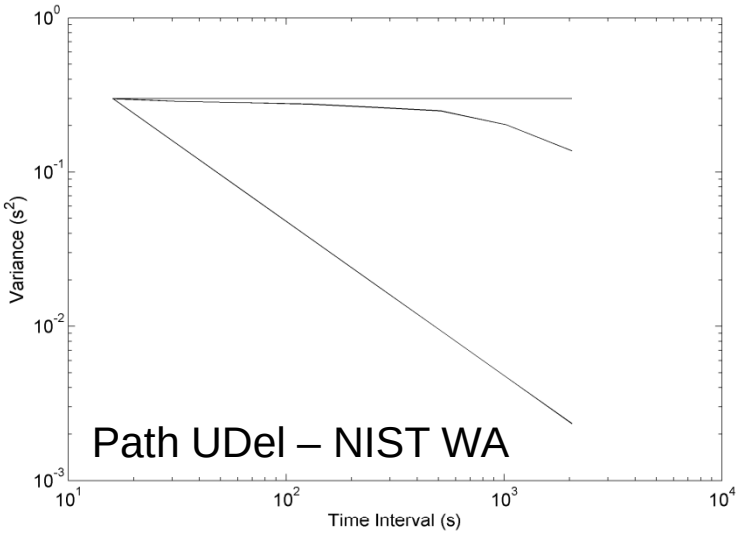
Timetelling in 1979

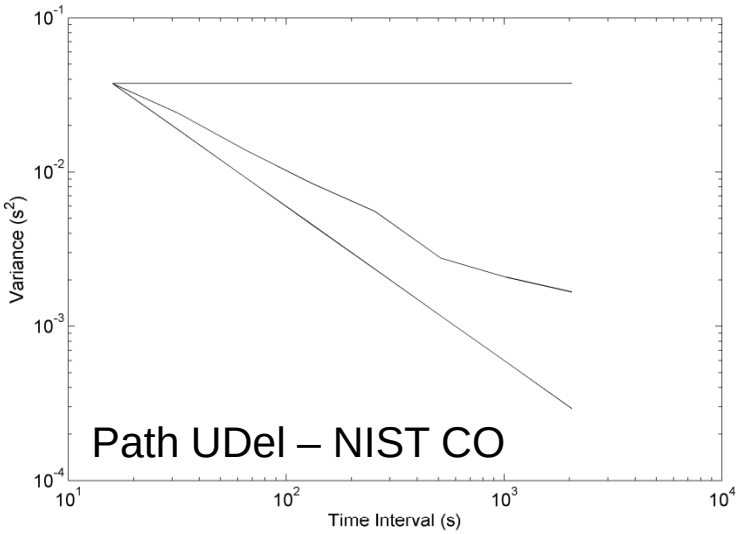


Timetelling in 2006



Long range dependency phenomena





- These variance-time plots demonstrate NTP reveals long-range dependency in the delay distribution.
- The slope -0.5 axis represents an exponential (memoryless) delay distribution; the horizontal axis represents random-walk
- The NIST CO plot suggests more of the former; the NIST WA path suggests more of the latter.
- These phenomena occur on paths all over the world.

A brief history of NTP time

- Time pull protocols are very old, like TIME and DAYTIME.
- NTP is almost as old, but it can both push and pull.
- A few highlights and fun along the way:

28 years and still ticking

- Time began in what became the Fuzzball *circa* 1979
  - Fuzzball hosts and gateways were synchronized using timestamps embedded in the Hello routing protocol
  - In 1981, four Spectracom WWVB receivers were deployed as Fuzzball primary reference. Two of these are still in regular operation, a third is a spare, the fourth was in the Boston Computer Museum
  - In 1982 the first version of NTP appeared in the Fuzzballs and was ported to Unix in 1984. It has evolved continuously since then.
- Timekeeping technology has evolved continuously over 20 years
  - Current NTP Version 4 improves performance, security and reliability, especially in congested network conditions.
  - Engineered Unix kernel modifications improve accuracy to the order of a few microseconds with precision sources.

Mommy, what's a Fuzzball?

- LSI-11 Fuzzballs were cloned in dozens of personal workstations, gateways and time servers in the US and Europe.
- Telnet, FTP, mail and other protocols were built and tested on these machines.
- On the left is the first Fuzzball, together with control box and 1200-bps modem. On the right is the last known Fuzzball, now in my basement.

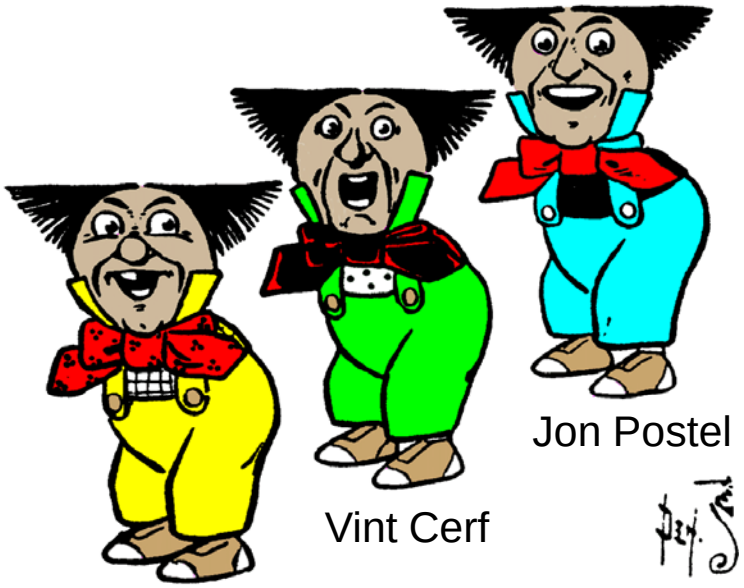
Evolution to NTP Version 4

- NTP Version 3 was first used in 1992 in a rambunctious Internet of congested links and 25-MHz workstations.
- NTP Version 4 architecture, protocol and algorithms have been evolved for gigabit networks and gigahertz computers.
  - Improved clock models which accurately predict the phase and frequency noise for each synchronization source and network path.
  - Engineered algorithms which reduce the impact of delay spikes and oscillator wander while speeding up initial convergence.
  - Redesigned clock discipline algorithm which can operate in frequency-lock, phase-lock and hybrid modes.
- For the ultimate performance, the clock discipline feedback loop has been implemented in the kernel for Solaris, Tru64, FreeBSD and Linux.

Lessons learned from NTP development

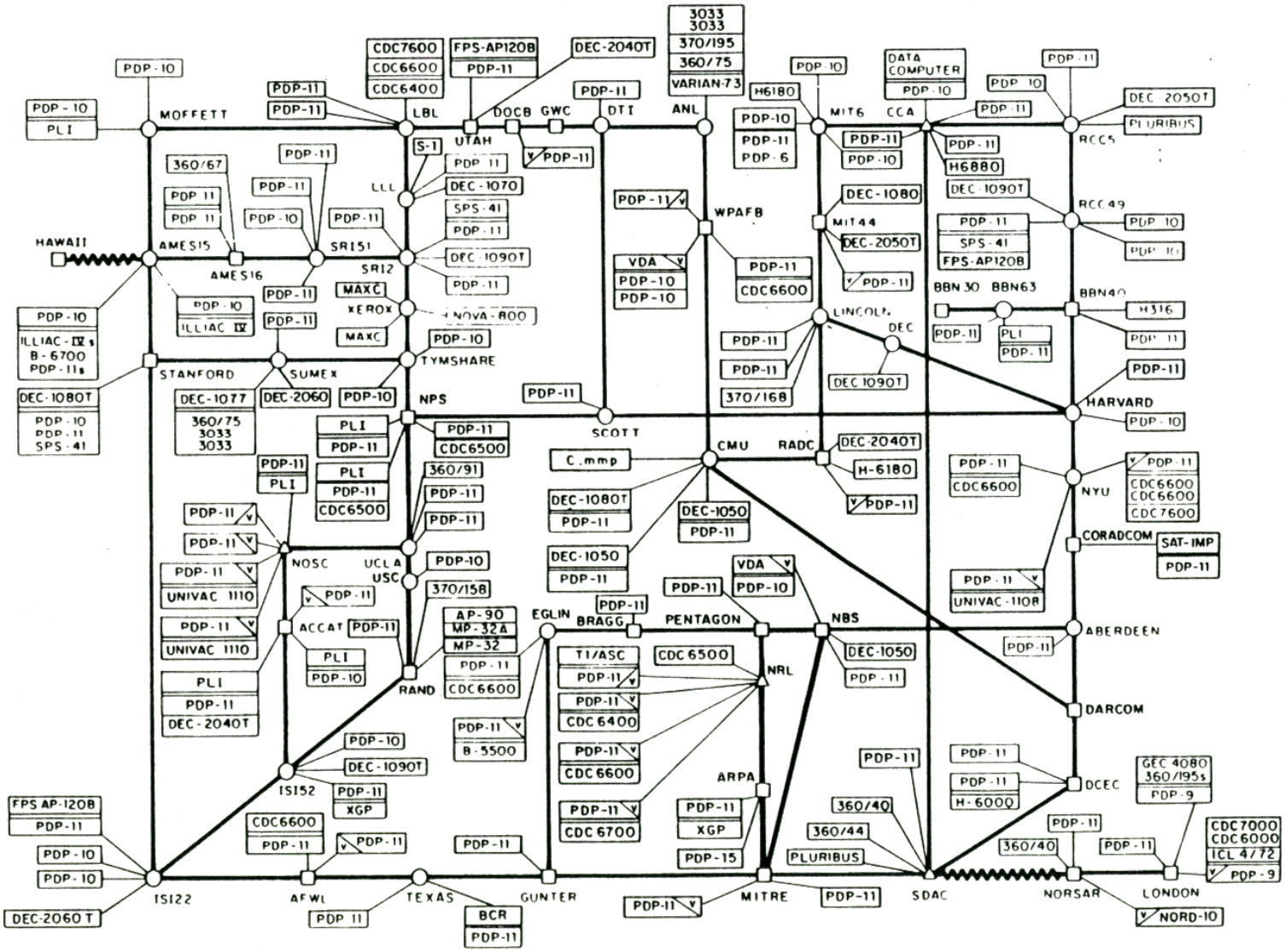
- Synchronizing global clocks with millisecond accuracy enables
  - the exact incidence of global events to be accurately determined.
  - Detection and prosecution of distributed denial of service attacks
  - real time synchronization of applications such as multimedia conferencing.
- Observations of time and frequency can reveal intricate behavior.
  - Usually, the first indication that some hardware or operating system component is misbehaving are synchronization wobbles.
  - NTP makes a good fire detector and air conditioning monitor by closely watching temperature-dependent system clock frequency.
  - Statistics collected in regular operation can reveal subtle network behavior and routing Byzantia.
  - NTP makes a good remote reachability monitor, since updates occur continuously at non-intrusive rates.

Internet paleontology

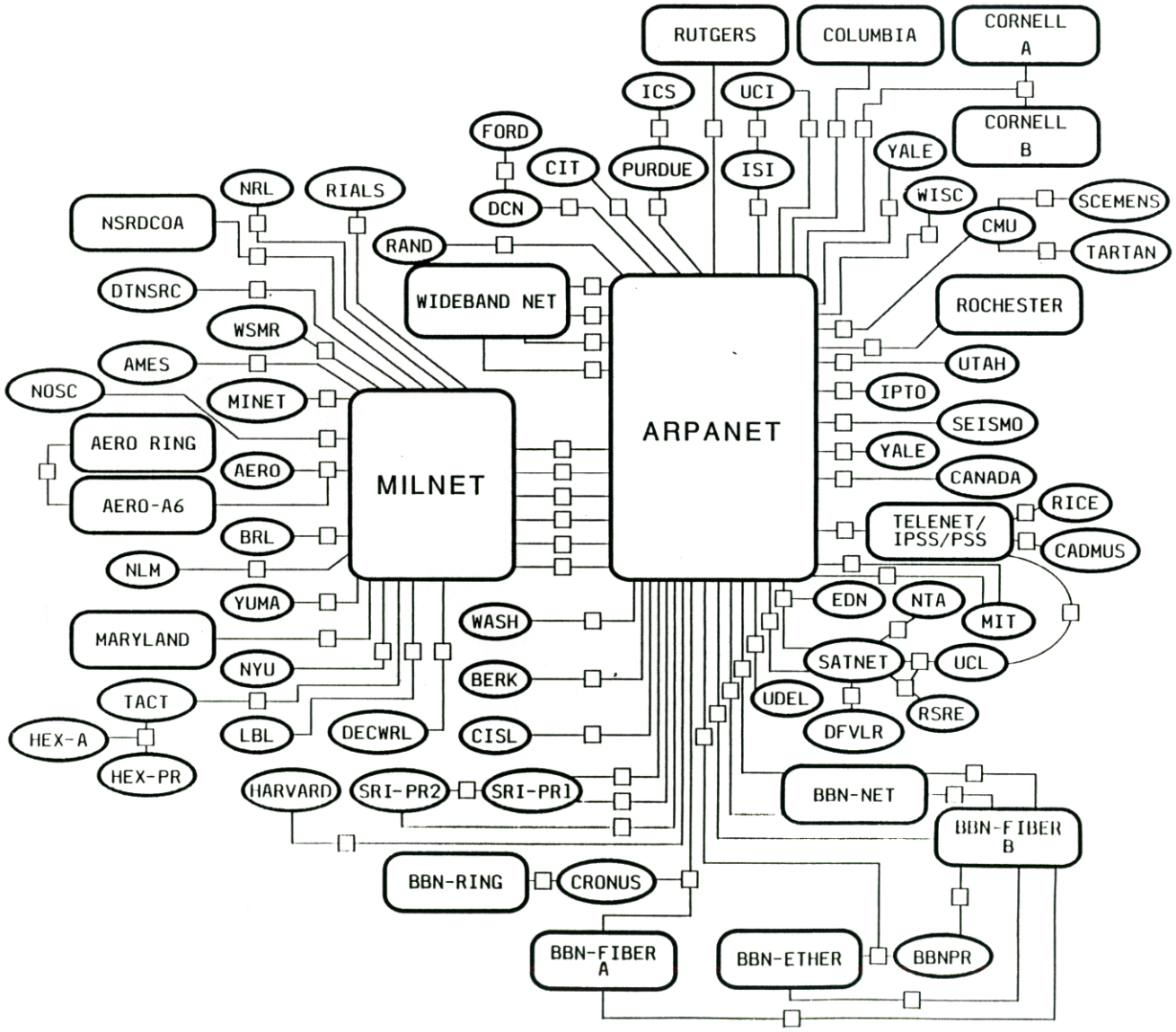


- ARPAcene (1976-1986): Evolution from ARPAnet terminal concentrator to NSF Internet backbone.
- NSFocene (1986-1996): Internet technology gets institutionalized, commercialized and politicized.
- Webocene (1996-2006): Everybody gets rich and famous.
- NTP grew and morphed as time went by.

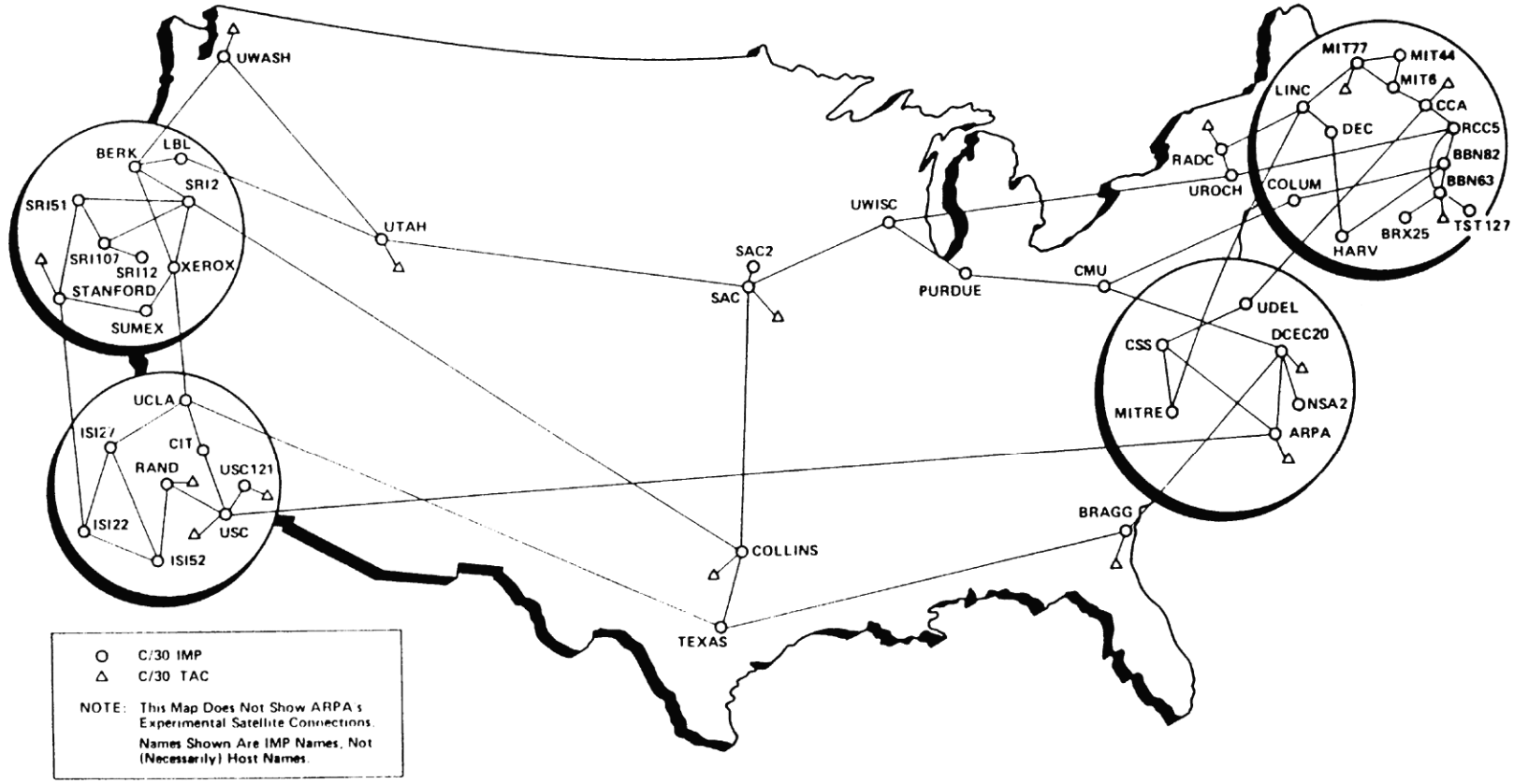
ARPAnet topology March 1979



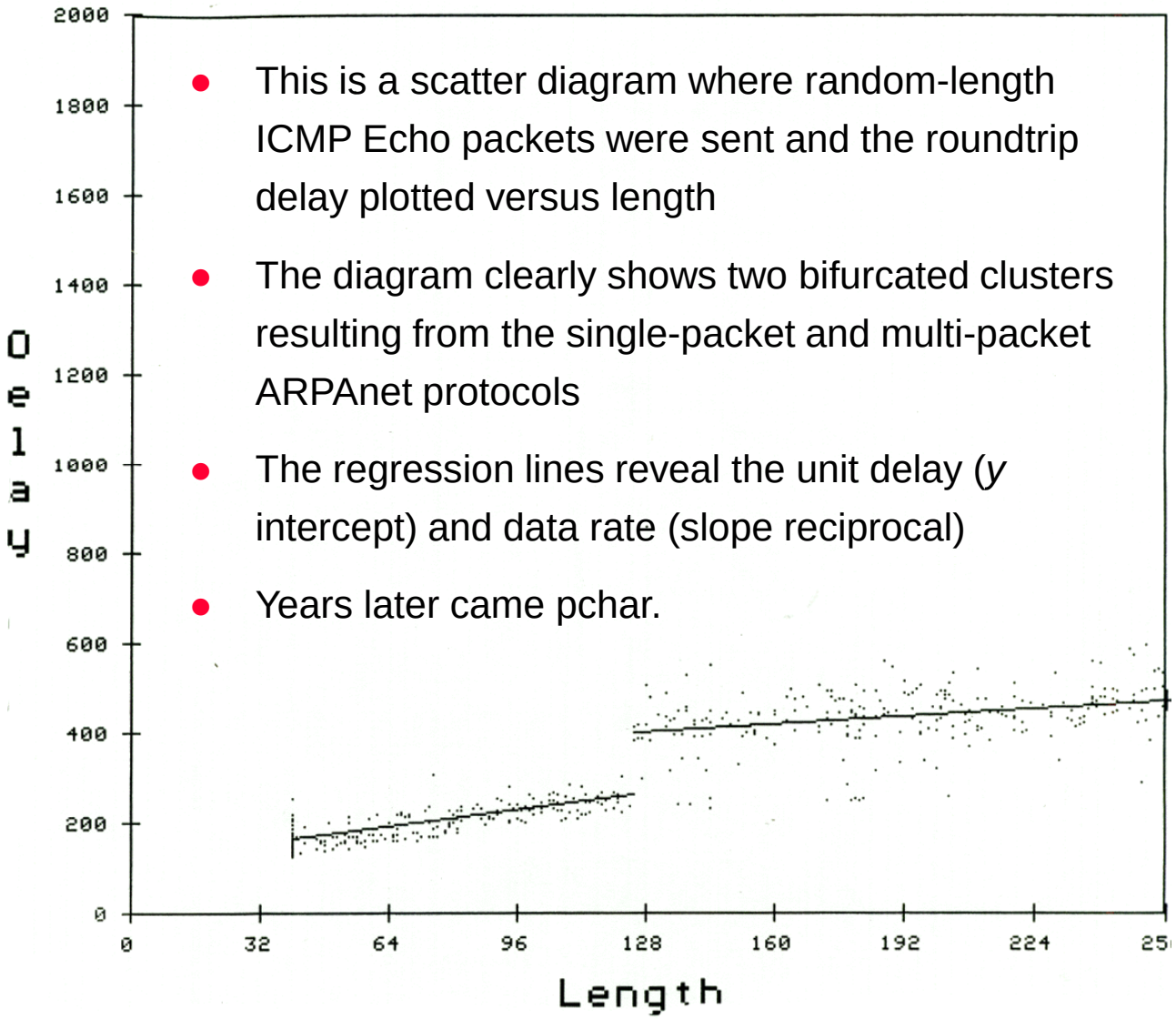
ARPANet/MILnet topology circa 1983



ARPAnet topology August 1986

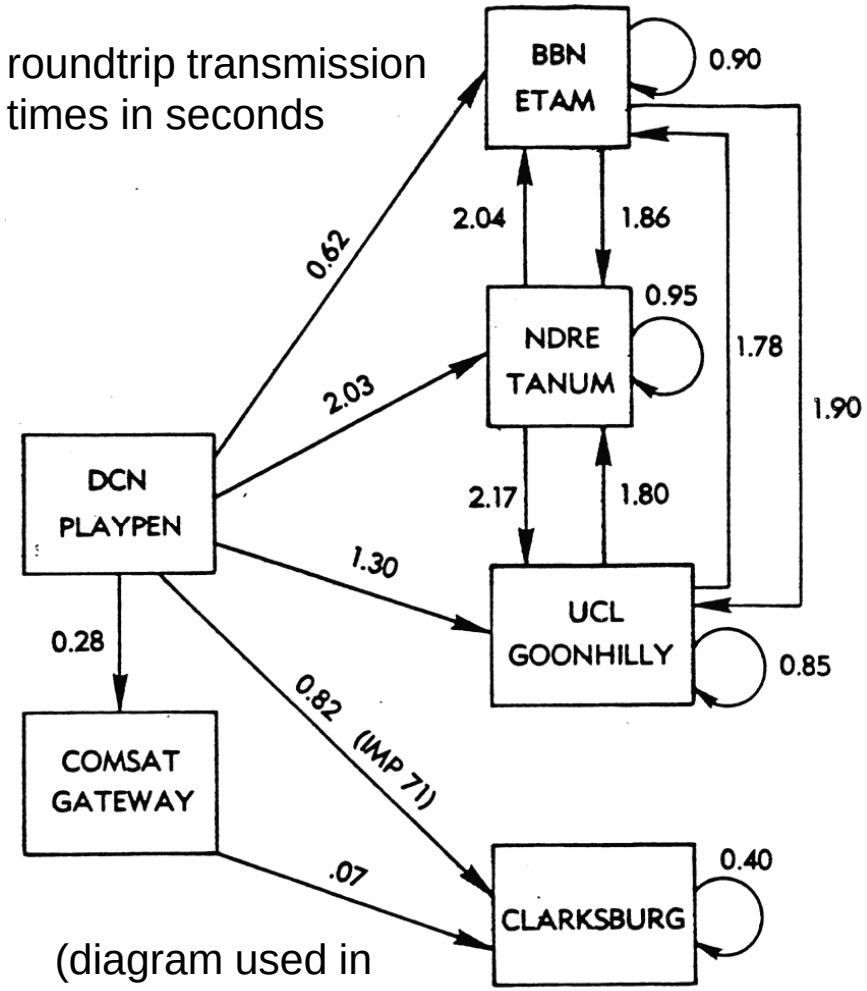


ARPAcene: the ancestor of NTP circa 1980



- This is a scatter diagram where random-length ICMP Echo packets were sent and the roundtrip delay plotted versus length
- The diagram clearly shows two bifurcated clusters resulting from the single-packet and multi-packet ARPAnet protocols
- The regression lines reveal the unit delay (*y* intercept) and data rate (slope reciprocal)
- Years later came pchar.

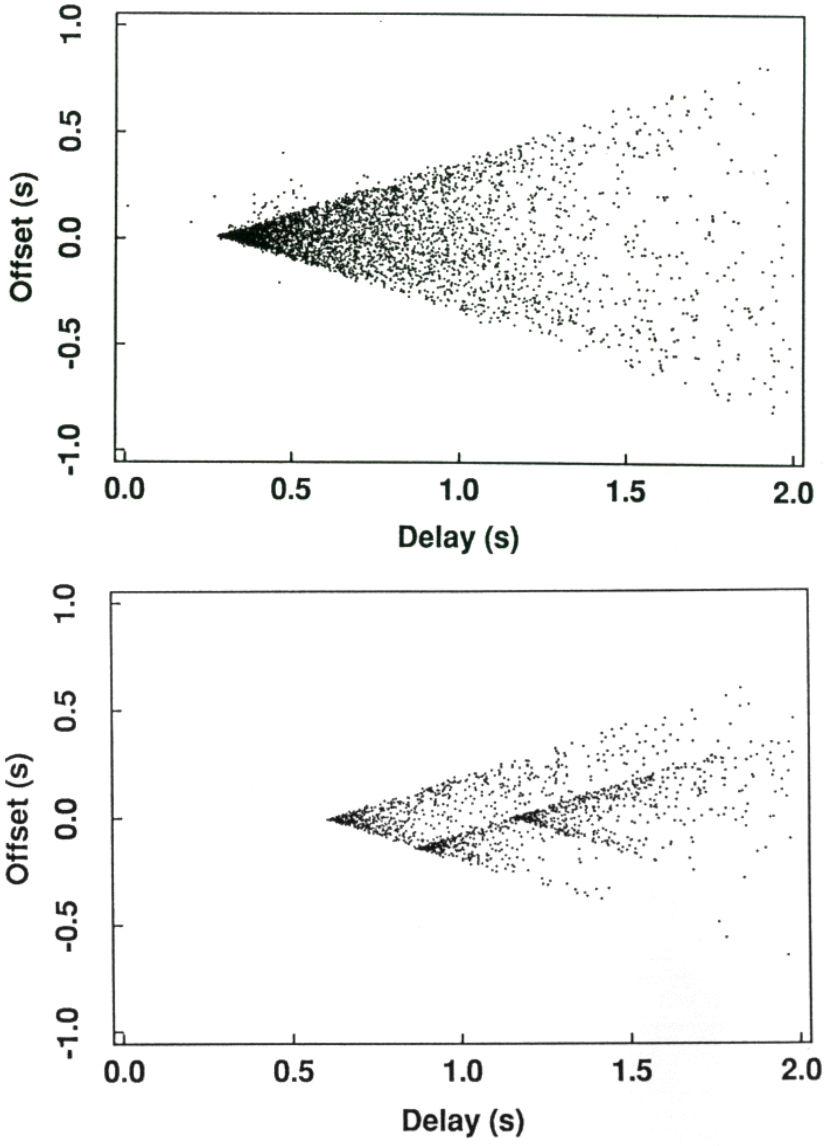
ARPAcene: SATnet measurement program



(diagram used in 1982 report)

- Earth stations in several countries were connected by a packet-switched INTELSAT satellite channel
- Stations supported scripted message generators and measurement tools
- Scripts were prepared transmitted via IP/TCP to experiment control program EXPAK, which ran in a designated ARPAnet host
- Once initiated, EXPAK launched the scripts and collected the results

NSFocene: network routing instability



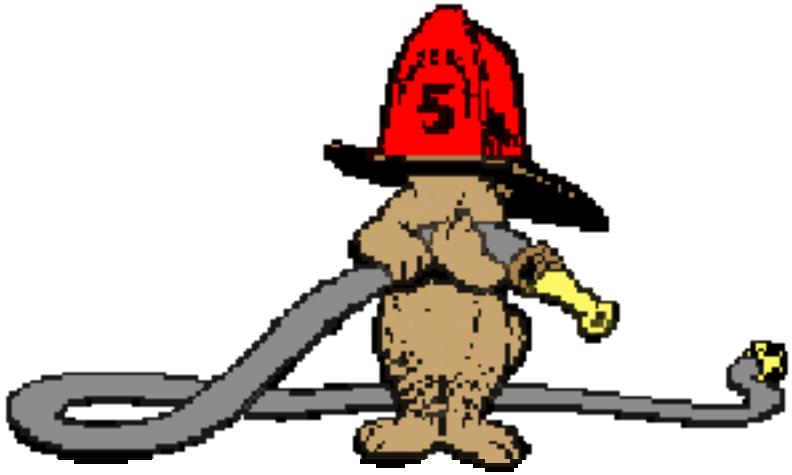
- These wedge scattergrams show the time offset plotted against delay for individual NTP measurements.
- For stable network routing, all points must lie within the wedge.
- The top diagram shows a typical characteristic with no route flapping
- The bottom diagram shows route flaps, in this case due to oscillation between landline and satellite links

NSFocene: IP/TCP measurements

TCP a fine mouthwash available in Britain

- While ARPAnet measurement tools had been highly developed, the Internet model forced many changes.
- The objects to be measured and the measurement tools could be in far away places like foreign countries.
- A number of distributed IP/TCP test, measurement and evaluation programs were conducted over the years.
- NTP was used in these programs to synchronize the clocks of the participants.
- NTP was also used in multimedia streaming applications and in a distributed orchestra demonstration.

Overload on the NTP public servers



- On the hazards of serving time.
- The U Wisconsin incident..
- Conditions at USNO
- Conditions at NIST.
- Things that can be done about it.

On the hazards of serving time

- With potential client populations in the millions, there is a very real vulnerability to grossly overload the public primary server population and/or the interconnecting networks.
- The public NTP client software exchanges packets with the server on a continuous low-rate basis in order to discipline the computer clock time and frequency.
- The sheer weight of numbers threatens to overwhelm at least some of the current NIST and USNO servers.
  - Other incidents reveal really bad network engineering and counterproductive
- Defective NTP client implementations have appeared that exhibit gross violations of the Internet social contract.
  - An example is the U Wisconsin incident reported in the next slide. parameter selection, especially poll interval.

The U Wisconsin incident

- U Wisconsin operates a number of time servers for campus access.
- A home router came on the market that
  - had the address of one of these servers hard-coded in firmware and could not be changed,
  - could send packets continuously at one-second intervals under conditions when the path or server was unavailable..
- This would not be a problem if only a small numbers of these routers were sold.
  - However, eventually 750,000 routers were sold and most could not be recalled, updated or even reliably found.
  - The resulting traffic overwhelmed the server, university network and service provider.
- There has been no wholly satisfactory solution to this problem other than to insure continuous service and to educate the manufacturer about socially responsible product design.

Conditions at USNO

- USNO operates about 20 NTP servers in the US, Alaska and Hawaii
  - Three of the busiest servers are in Washington, DC.
  - They share an aggregate load of 3,000-7,000 packets per second (PPS).
  - While these three servers are at only 10 percent capacity, the network is badly overloaded, leading to significant packet loss and badly degraded time quality.
- Much of the traffic is from clients sending at unrealistic rates.
  - In one case the client is spraying at 14 PPS, a rate equivalent to 731 properly configured NTP clients.
  - In another case a university firewall has channeled 2,000 campus clients separately to the USNO servers when it should synchronize to USNO and have all campus clients synchronize to it, as NTP is designed to do.

Conditions at NIST

- NIST operates about a dozen NTP public time servers in the US.
  - Three of the busiest servers are in Boulder, CO
  - They share an aggregate load similar to USNO, but the NIST network infrastructure is far more resilient than USNO.
- An experiment collected statistics in a nine-second window on each machine using a sampling technique which captured about 13 percent of the arrivals.
  - The results revealed over 500 clients with polling intervals of 5 seconds or less and 15 with poll intervals less than one second. Well behaved NTP clients send at rates usually at intervals of fifteen minutes or more.
  - Most incidences involved packet bursts lasting from a few seconds to several days and separated by minutes, hours or days.
  - One particularly offensive elephant was sending continuously at two packets per second.

Things that can be done about it

- Some things are obvious
  - Rig the host name/address translation (DNS) to lie, cheat and steal; that is, randomize the addresses over a geographically dispersed server population (e.g., NTP pool scheme).
  - Find ways to deflect traffic from congested servers (e.g., time.nist.gov) to less busy servers closer to users (BGP Anycast).
  - Never ever carve an unconfigurable server address in the firmware.
- Educate potential implementors about best and worst practices in the selection of protocol parameters, especially poll interval.
  - RFC 4330 includes clearly defined best practices which minimize network and server.
  - Boil best practices volators in oil. Smelly, stinky, public-ridicule oil.
- More aggressive pro-active nastiness may be necessary.
  - The Kiss-o'-Death packet is designed to disarm clients as necessary.
  - The Call-Gap scheme finds the elephants and shoots them.

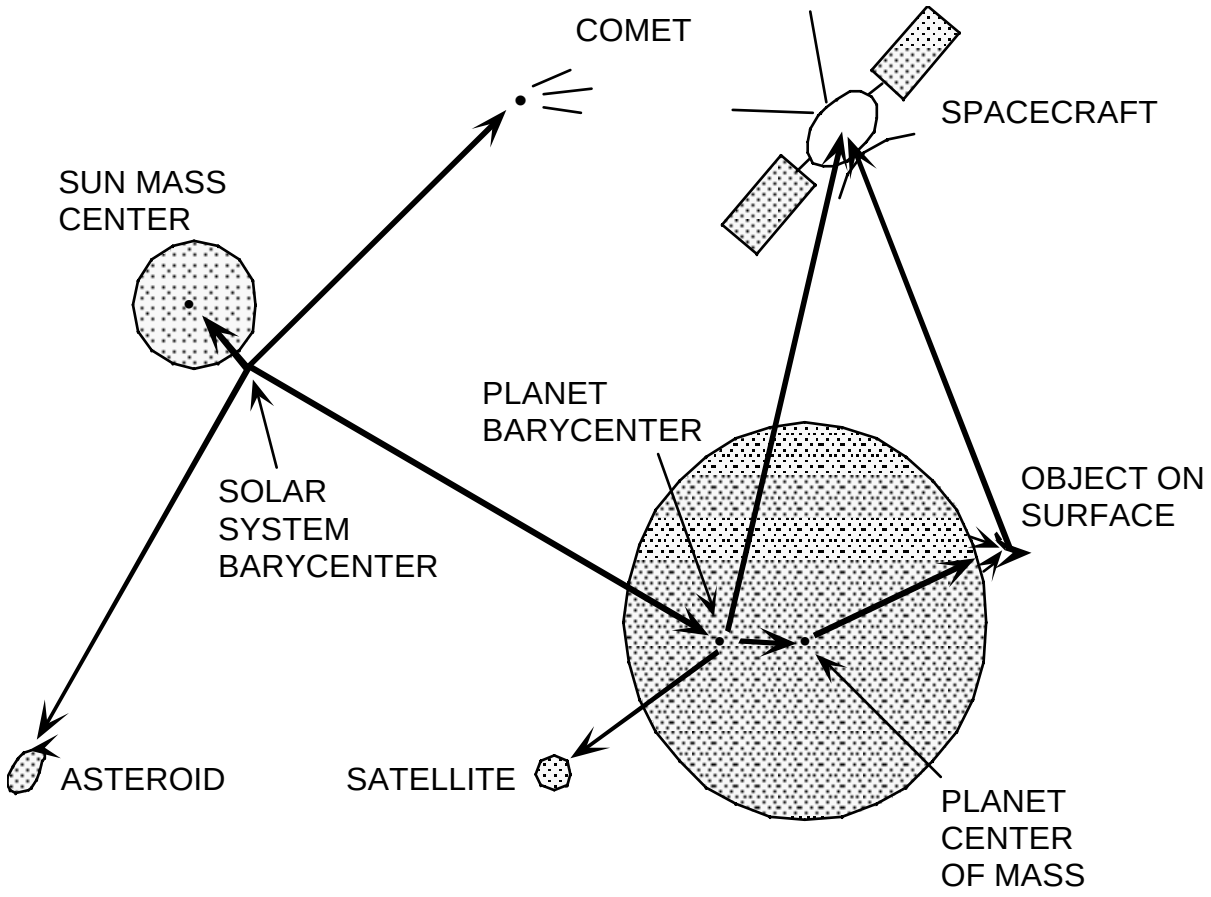
NTP and the interplanetary internet

- NASA and JPL are extending the Internet for space exploration.
- Due to long propagation times, existing Internet protocols like TCP and NTP are unsuitable.
- However, NTP can be extended with an iterative algorithm that does the messy celestial mechanics before handing off to NTP.

The Interplanetary Internet (IPIN)

- Research program funded by DARPA and NASA.
- Near term emphasis on Mars exploration and mission support.
  - Surface base stations and rovers – perform experiments, collect data.
  - Satellite orbiters – relay commands to base stations, retrieve data for later transmission to Earth.
  - Spacecraft – transports orbiters and surface vehicles to Mars.
- The Interplanetary Internet (DNS .sol)
  - NASA Deep Space Network (DSN) – three huge antenna farms in California, Spain and Australia, time shared for Mars and other NASA missions
  - Earth internet – coordinate mission activities, send commands and retrieve data via DSN, disseminate results
  - MARS internet – communicate between DSN, orbiters and surface vehicles; perform housekeeping functions such as antenna pointing, network routing, ephemeris maintenance and general timekeeping

Space and planetary vehicles



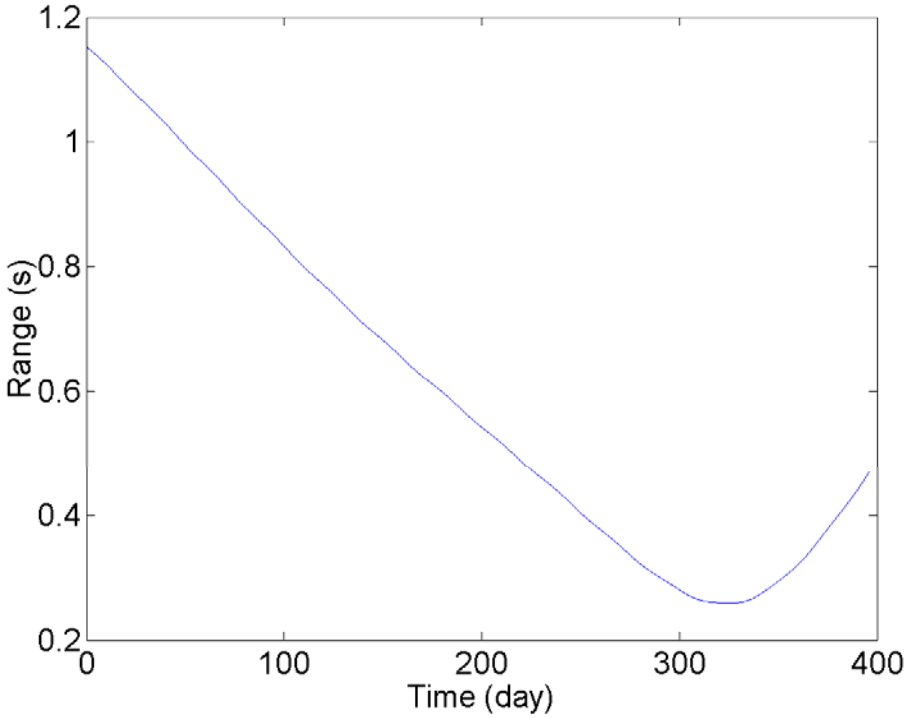
IPIN communication issues

- Transmission delays between Earth and Mars are variable and in general much longer than in Earth and Mars internets.
- Transmission speeds are asymmetric and highly variable and in general far slower than Earth internet.
- Connectivity between Mars surface and orbiters and between Earth and Mars is not continuous, but opportunities can be predicted.
- Error recovery using retransmissions is impractical; TCP is useful only in Earth internet and Mars internet, but not between Earth and Mars
- Dependency on Earth-based databases is not practical on Mars, so any databases required must be on or near Mars

NTP time and position algorithm

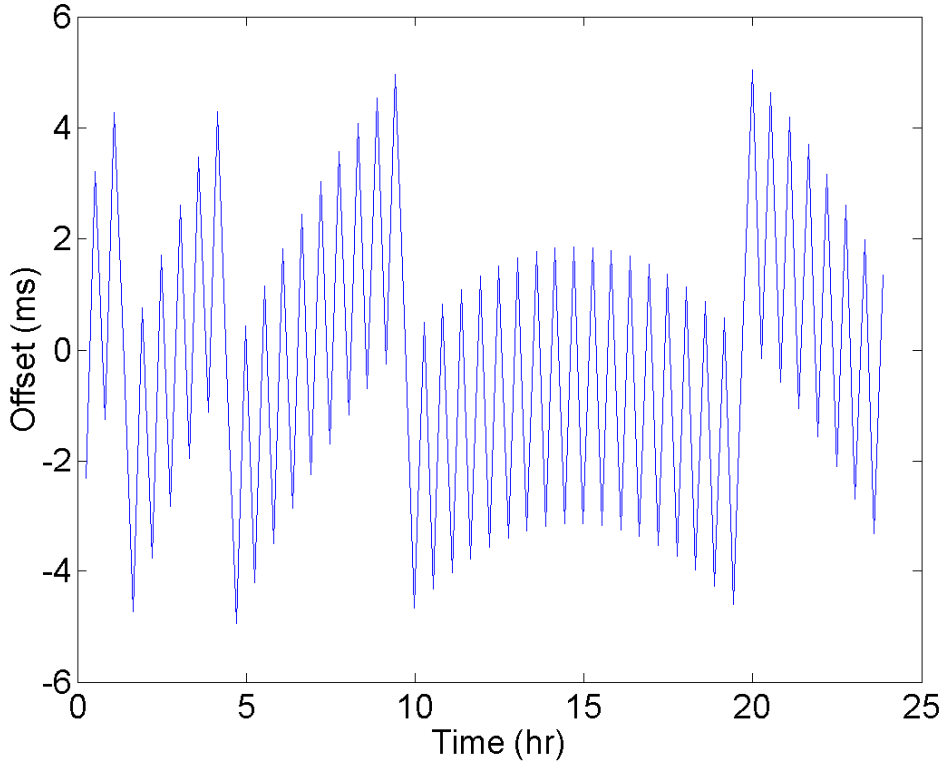
- All clients and servers operate on UTC timescale for compatibility with current Earth Internet.
- Servers and clients have onboard ephemeris and SPICE routines to calculate position at any given UTC time.
- A server sends a transmit timestamp and computed ephemeris position according to its clock.
- The client corrects its clock using an iterative procedure.
  - The client records the apparent receive timestamp and computes the ephemeris position according to its clock.
  - It then calculates the actual receive timestamp using the calculated free-space propagation time (lighttime).
  - It corrects its clock and computes a new position based on the corrected clock, then calculates a new lighttime and receive timestamp.
  - The client iterates this procedure until the corrections converge.

Earth-Mars lighttime range



- Earth-Mars lighttime range (s) over some 400 days calculated from ephemerides.
- This is used to compensate for the delay before handing off to the conventional NTP algorithms.

Earth-Mars prediction error



- Residual jitter measured by NTP after lighttime compensation
- This may be due to Chebyshev interpolation residuals.

NTP security protocol

- NTP is not like other protocols that depend on a reliable time infrastructure. NTP *is* the time infrastructure. So, how do we reliably authenticate NTP servers to dependent clients?
- NTP can’t use the conventional PKI infrastructure, as it can’t trust signatures unless the period of validity can be verified.
- NTP has to build its own infrastructure using timestamped certificates and crafted zero-knowledge proofs.

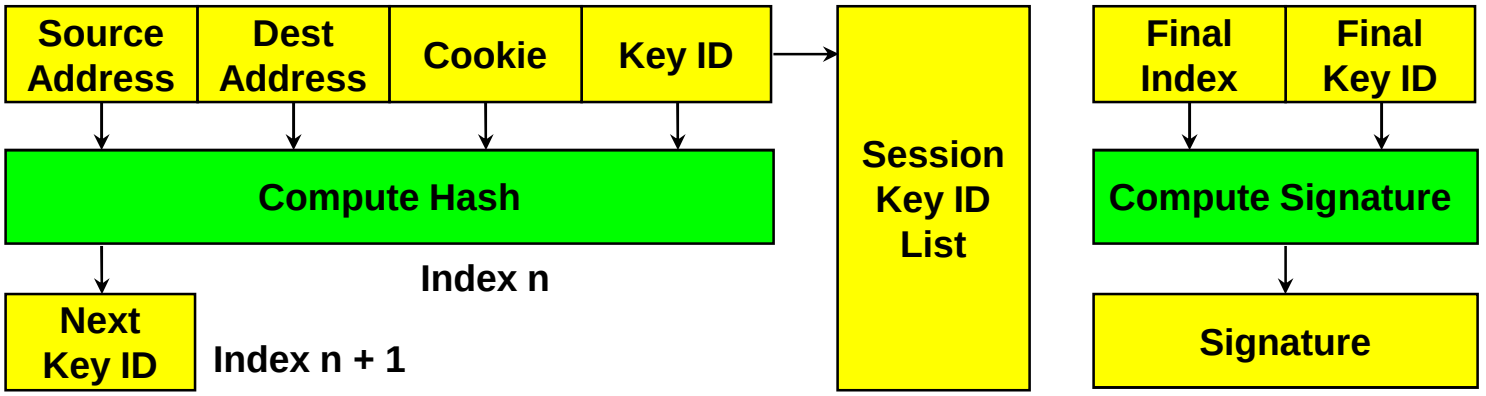
NTP security model

- It authenticates the server to the client, not the other way around.
- It must be based on public key cryptography using Internet PKI and standard certificates with verified period of validity.
- It must interoperate with existing NTP protocol modes (including broadcast) and symmetric key cryptography.
- It must provide for the independent collection of cryptographic values and time values.
- It must not significantly degrade the potential accuracy of the NTP synchronization algorithms.
- It must be resistant to cryptographic attacks. It must tolerate occasional lost, duplicate or out of order packets.
- All cryptographic algorithms must be obtained from public cryptographic libraries (OpenSSL).

Autokey security protocol

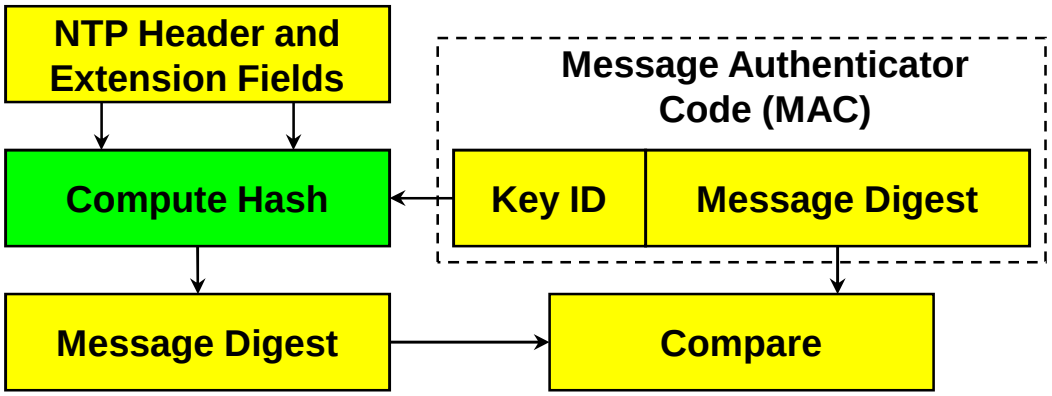
- NTP and Autokey protocols work independently using the same packets.
- An NTP server synchronizes to servers at the next lower stratum level. At the bottom (stratum 1) level are the trusted primary servers.
- The protocol uses an exchange of signed messages to obtain and verify certificates and other credentials.
- The server constructs a certificate trail from a trusted server using public keys augmented by zero-knowledge proofs (challenge/response exchange).
- When server time and at least one certificate trail are verified, the host is admitted to the population used to synchronize the system clock. Subsequent packets are not signed.
- Subsequent NTP messages are individually authenticated using a session key and message digest (keyed MD5).

Sending messages



- The initial session key is constructed using the addresses, private cookie and a random initial key ID.
- Subsequent session keys are constructed using the first four octets of the previous session key as the new key ID.
- The final index number and last key ID are provided to the client in a signed message.
- The server uses the session key ID list in reverse order and discards each key after use.
- When the list is exhausted, a new list is generated.

Receiving messages



- The client first verifies the hash of the current key ID matches the last one received. In case of loss, it might need to do this again, but not more than the previously provided final index.
- The client reconstructs the session key and key itself, then verifies the result matches the message digest .

Miscellaneous photos and references

- Gadget box PPS interface
- Used to interface PPS signals from GPS receiver or cesium oscillator
  - Pulse generator and level converter from PPS signal edge
  - Simulates serial port character or stimulates modem control lead

Udel Master Time Facility (UMTF)

Spectracom 8170 WWVB Receiver

Spectracom 8183 GPS Receiver

Spectracom 8170 WWVB Receiver

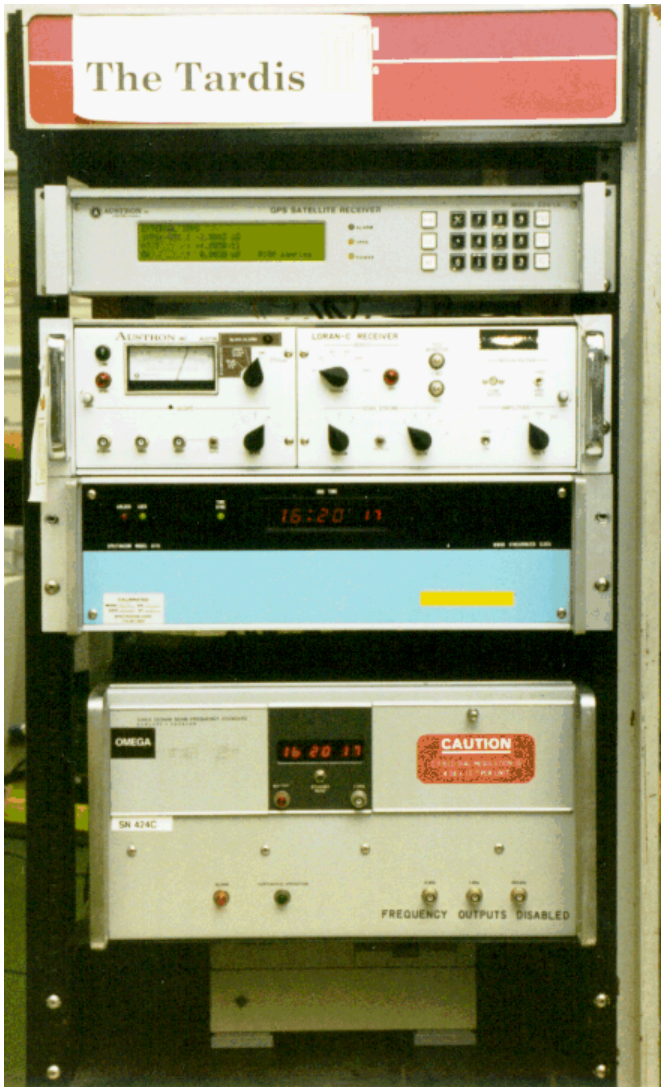
Spectracom 8183 GPS Receiver

Hewlett Packard 105A Quartz Frequency Standard

Hewlett Packard 5061A Cesium Beam Frequency Standard

NTP primary time servers *rackety* and *pogo* (elsewhere)

Precision timekeeping equipment



Austron 2200A GPS Receiver

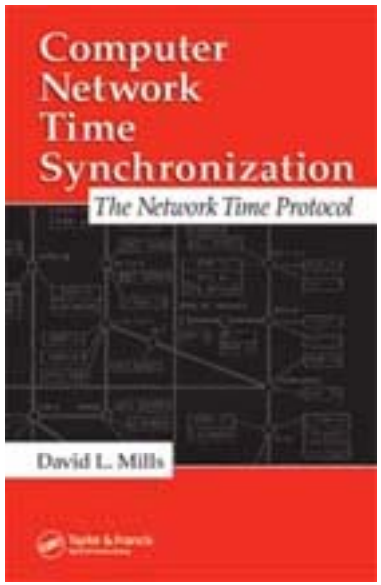
Austron 2000 LORAN-C Receiver

Spectracom 8170 WWVB Reciver

Hewlett Packard 5061A Cesium Beam Frequency Standard

NTP primary time server *rackety*

Further information



- Mills, D.L. *Internet Time Synchronization – the Network Time Protocol*. CRC Press, 2006, 304 pp.
- Network Time Protocol (NTP): http://www.ntp.org/
  - Current NTP software, documentation.
  - FAQ and links to other sources and interesting places.
- NTP Project page: http://www.eecis.udel.edu/~mills/ntp.html
  - Papers, reports and memoranda in PDF format
  - Briefings in PowerPoint and PDF formats
  - Collaboration resources hardware, software and documentation