Latenz

Wenn eine Reaktion auf sich warten lässt – Teil 1

Im täglichen Leben begegnen wir häufig Latenzen, ohne uns dessen bewusst zu sein. Gesellschafts- und Naturwissenschaften kennen vielfältige Ausprägungen dieses Begriffs und auch in der Kommunikationstechnik spielt er eine erhebliche Rolle. Doch was muss man sich unter Latenz vorstellen? Die Herkunft des Latenzbegriffs ist im lateinischen Adjektiv „latens" zu finden, das die Bedeutung von heimlich oder verborgen hat. Das Substantiv Latenz beschreibt demnach die Existenz einer versteckten und somit (noch) nicht sichtbaren Sache, Wirkung oder Eigenschaft.

Die Beispiele für Latenzen sind vielfältig. In der Medizin spricht man von einer latenten Infektion oder Krankheit, wenn deren Auswirkungen noch nicht in Erscheinung getreten sind, also bis zum Ausbruch noch „schlummern". Ein Mensch, der stets mit seinem Schicksal hadert, wird in der Psychologie als latent unzufrieden bezeichnet. In der Soziologie wird einem Teil der Bevölkerung latenter Rassismus bescheinigt und in der Politik kennt man den Begriff der latent schwelenden Krisen. Der Physiker spricht von latenter Wärme, wenn die von einer Masse aufgenommene (oder abgegebene) Wärmeenergie nicht von einem proportionalen Temperaturanstieg (oder -rückgang) begleitet wird. Ein klassisches Beispiel dafür ist der Übertritt von Wasser aus der flüssigen in die feste Phase (Eis). Geschieht dies durch kontinuierlichen Entzug von Wärmeenergie, so ist damit ein ebenso kontinuierlicher Temperaturabfall festzustellen. Beim Erreichen des Gefrierpunkts (0 °C) stagniert die Temperatur allerdings so lange, bis weitere 80 kcal/kg entzogen wurden, um dann weiter proportional abzufallen. Der Grund dafür ist, dass die bei dieser Temperatur entzogene Wärmeenergie (Latenzwärme) den Wechsel von der flüssigen in die feste Phase bei gleichbleibender Temperatur bewirkt. Mit anderen Worten: Es werden die Bindungskräfte zwischen den Molekülen verändert, ohne deren kinetische Energie zu beeinflussen. Entsprechendes gilt beim Übergang zwischen flüssiger und dampfförmiger Phase.

Die allgemein verzögernde Wirkung von Latenz führt meist zu unerwünschten Effekten, weil sie die Reaktion auf eine Anregung während einer häufig nur ungenau vorhersehbaren Zeitdauer aussetzt. Besonders in der Regelungs- und Kommunikationstechnik spielen Latenzen eine wichtige Rolle. Oft werden hier allerdings nicht völlig deckungsgleiche Synonyme wie Laufzeit, Totzeit, Reaktionszeit, Verweilzeit, Verarbeitungsdauer usw. verwendet.

Latenzen in der Kommunikationstechnik

Prinzipiell kann die Reaktion auf eine Anregung erst mit einer gewissen Verzögerung erfolgen. Der Grund darin ist in der endlichen Lichtgeschwindigkeit von 300.000 km/s zu suchen, die nicht überschritten werden kann. Sie ist der Übermittlung einer Anregung von der Nachrichtenquelle zur Nachrichtensenke, deren Verarbeitungszeit und der Rückleitungszeit der Reaktion zur Quelle mindestens zugrunde zu legen. Bei der kommunikationstechnischen Überwindung großer Distanzen wird die Reaktionszeit von den darin enthaltenen Laufzeitanteilen dominiert. In komplexen digitalen Systemen kommen die Zeitbedarfe für die Signalverarbeitungsschritte A/D-Wandlung, Abarbeitung der Algorithmen, Speicherung, Pufferung, Synchronisation, D/A-Wandlung, Modulation usw. hinzu.

Weltraumfahrt

Ein gutes Beispiel für laufzeitdominierte Latenz ist die Raumsonde Voyager 1, die am 5. September 1977 vom amerikanischen Weltraumbahnhof Cape Canaveral gestartet wurde (Bild 1). Zu ihr besteht nach über 40 Jahren Missionsdauer immer noch eine Kommunikationsverbindung über eine Distanz von inzwischen 141,6 Millionen Astronomischen Einheiten (1 AU = 149.600.000 km), entsprechend 21,1 Milliarden Kilometer. Die sich mit Lichtgeschwindigkeit ausbreitenden Funksignale sind also auf Hin- und Rückweg jeweils 19 Stunden und 36 Minuten unterwegs. Als am 29. November 2017 einige Schubdüsen aktiviert wurden (was zuletzt 37 Jahre vorher im November 1980 in der Nähe des Saturn geschah), um eine Antenne besser in Richtung Erde auszurichten, mussten die Wissenschaftler quälende 38 Stunden warten, bis sie sich vom Erfolg dieser Maßnahme überzeugen konnten. Jetzt erhofft sich die NASA, die Lebensdauer der Sonde um zwei bis drei Jahre verlängert zu haben, ehe die Radionukleidbatterie an Bord erschöpft ist.

Bild 1: Seit 40 Jahren unterwegs: Die Raumsonde Voyager 1 ist knapp 20 Lichtstunden von der Erde entfernt.

Um die Weiten des Weltalls zu erahnen, sollte man sich vor Augen führen, dass es noch 38.000 Jahre dauern wird, bis Voyager 1 den nächsten Stern mit der Bezeichnung AC+793888 (Gliese) im Sternbild Kleiner Bär (Kleiner Wagen) passiert. Dann beträgt die Erdentfernung 17,6 Lichtjahre. Die Schwanzspitze des Kleinen Bärs (oder das Deichselende des Kleinen Wagens) ist übrigens der Polarstern mit einer Erdentfernung von knapp 434 Lichtjahren.

Man kann sich gut vorstellen, dass es völlig unmöglich wäre, von der Erde aus ein derart entferntes interstellares Flugobjekt durch ein dichtes kosmisches Trümmerfeld zu manövrieren, wo eventuell Minuten oder wenige Stunden über eine Kollision entscheiden. Hier stehen die erforderliche und die tatsächliche Latenz der zur Navigation erforderlichen Steuerschleife in einem unlösbaren Missverhältnis. So wird auch die Kommunikation der Menschheit mit fremden Intelligenzen aus den Tiefen des Weltalls weit außerhalb unseres Sonnensystems Science-Fiction bleiben. Wird es uns in Millionen von Jahren überhaupt noch geben?

Bild 2: In den Anfangszeiten der Telefonie lagen die Laufzeiten des Sprachsignals bei wenigen Millisekunden.

Von Mensch zu Mensch

In der zwischenmenschlichen Kommunikation werden Latenzzeiten (Einweg-Laufzeit Mund → Ohr) ab einer gewissen Länge als störend empfunden.

In den Anfangszeiten der Telefonie wurden Latenzen im Gesprächsverlauf im Wesentlichen durch die Signallaufzeiten im Kabel (typ. 200.000 km/s) (Bild 2) verursacht. Diese lagen in Verbindungen zwischen den deutschen Großstädten bei wenigen Millisekunden und waren nicht wahrnehmbar.

Bei der digitalen Telefonie gilt als obere Akzeptanzgrenze nach ITU-T-Empfehlung G.114 eine Latenzzeit von 400 ms. Ein hoch interaktiver sprachlicher Informationsaustausch kann aber durch Latenzen von mehr als 100 ms bereits behindert werden. Der Grund liegt darin, dass der Angesprochene eine Zeitlang warten muss, bis der Beitrag seines Gesprächspartners bei ihm eintrifft, er dann eine kleine Denkpause einlegen muss, um seine Antwort auf den Weg zu bringen, und diese wiederum eine gewisse Zeit für den Rücklauf benötigt. Diese Verzögerung kann beim jeweils anderen Telefonpartner den Eindruck entstehen lassen, er sei nicht gehört worden. Wenn er deshalb eine Nachfrage innerhalb der Latenzzeitspanne abschickt, kollidiert diese gewissermaßen mit der eigentlich erwünschten und inzwischen eintreffenden Reaktion. Man fällt sich ständig ins Wort und eine harmonische, von Spontaneität gekennzeichnete Konversation ist praktisch nicht möglich. Damit wird verständlich, warum Telefonie über Satelliten im geostationären Orbit problematisch ist.

An einem Beispiel soll dies verdeutlicht werden. Betrachten wir ein auf dem Internetprotokoll (IP) basierendes Telefonat über eine transatlantische Satellitenverbindung (Bild 3). Teilnehmer 1 (USA) ruft Teilnehmer 2 (Deutschland) an (rote Pfeile). Der Ruf wird digitalisiert und geht über eine Festleitung an die Satellitenbodenstation 1. Bis zu ihrem Eintreffen dort mögen 50 ms vergangen sein. Nach der Aufbereitung zum sendefähigen Signal (50 ms) erfolgt der Uplink zum etwa 40.000 km entfernt über dem Atlantik stehenden geostationären Satelliten (120 ms). Hier wird das Signal für den Downlink nach Europa konditioniert (100 ms) und trifft 120 ms später bei Satellitenbodenstation 2 ein. Hier wird es demoduliert und in das Festnetz eingespeist (50 ms). Nach abermals 50 ms Laufzeit trifft es als reanalogisiertes Schallsignal am Ohr von Teilnehmer 2 ein. Bis der Anfang der Nachricht die Strecke vom Mund des Anrufers bis zum Ohr des Angerufenen (mouth to ear) zurückgelegt hat, sind also 540 ms vergangen. Bei einer angenommenen Symmetrie des Rückwegs müssen daher abermals 540 ms vergehen, bis der Beginn der Antwort des Angerufenen am Ohr des Anrufers eingeht. Alles in allem muss der Anrufer also mindestens 1,08 s abwarten, bevor er mit dem Eingang einer Antwort des Angerufenen rechnen kann (Antwortzeit). Für ein zwar verlangsamtes, aber halbwegs flüssiges Gespräch ist deshalb eine hohe Disziplin der beiden Gesprächspartner zum Abwarten der Latenzzeiten erforderlich. Der Ersatz der Satellitenstrecken durch eine 10.000 km lange Glasfaserverbindung (Seekabel) mit einer Signalgeschwindigkeit von 2/3 der Lichtgeschwindigkeit würde die Antwortzeit auf etwa 150 ms verkürzen.

Der Ablauf eines fiktiven minimalistischen Gesprächsablaufs ist in Bild 4 dargestellt. Ohne Latenzzeiten wäre das Telefonat nur 11 s (100 %) lang. Die Latenzen verlängern es um 2,7 s (24,55 %) auf 13,7 s (124,55 %). Für die Übertragung des gleichen Informationsgehalts wird also mehr Zeit benötigt, was einem Rückgang der Übertragungsgeschwindigkeit um etwa 20 % entspricht. Dieses einfache Beispiel macht also plausibel, dass Latenzen im Verlauf einer Kommunikation die durchschnittliche Datenrate reduzieren und deshalb im Rahmen des physikalisch Möglichen zu minimieren sind.

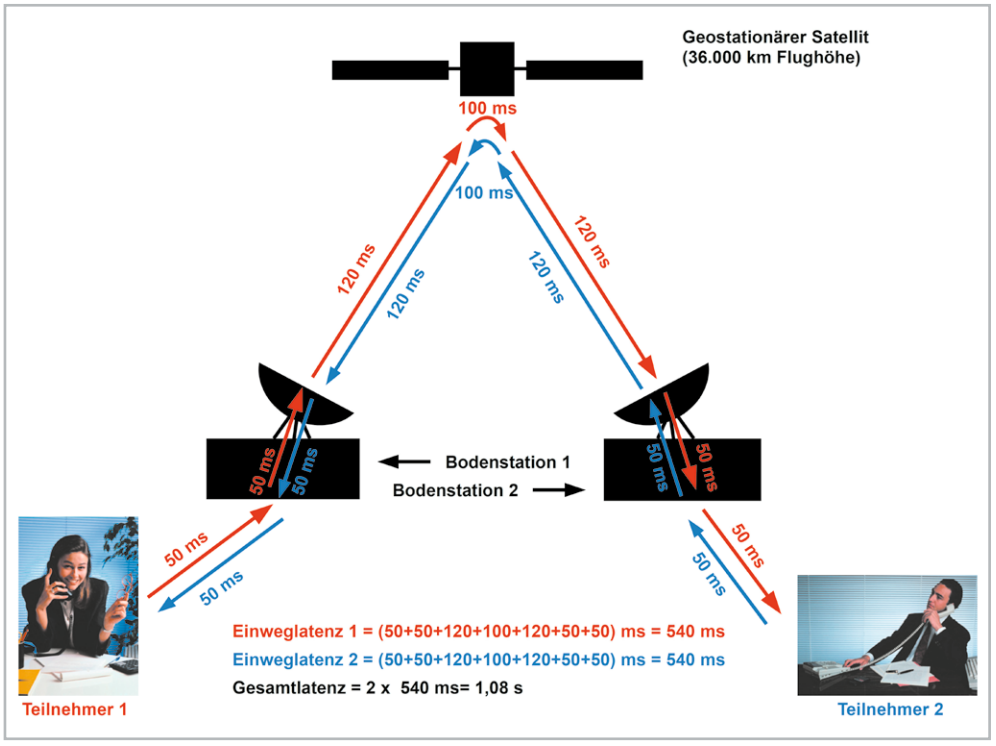

Bild 3: Bei Telefonverbindungen über einen geostationären Satelliten treten deutlich störende Latenzzeiten auf. Ein flüssiges, spontanes Gespräch wird dadurch praktisch unmöglich.

Bis jetzt haben wir nur die Laufzeit eines Signals als Ursache für Latenzen betrachtet. Doch auch die Zeit, die zur Verarbeitung einer Information benötigt wird, trägt zur Latenz bei. Sie kann durch die Komplexität der zu verarbeitenden Information oder die Leistungsfähigkeit der informationsverarbeitenden Technik bestimmt sein. Auf jeden Fall bewirkt die Verarbeitungszeit ein verzögertes Eintreffen der Antwort beim Kommunikationspartner und ist für ihn gar nicht von einer physikalischen Laufzeit zu unterscheiden. Dazu ein Beispiel, das vielleicht zum Schmunzeln einlädt, aber das Gesagte illustriert. Ein Gesprächspartner, der

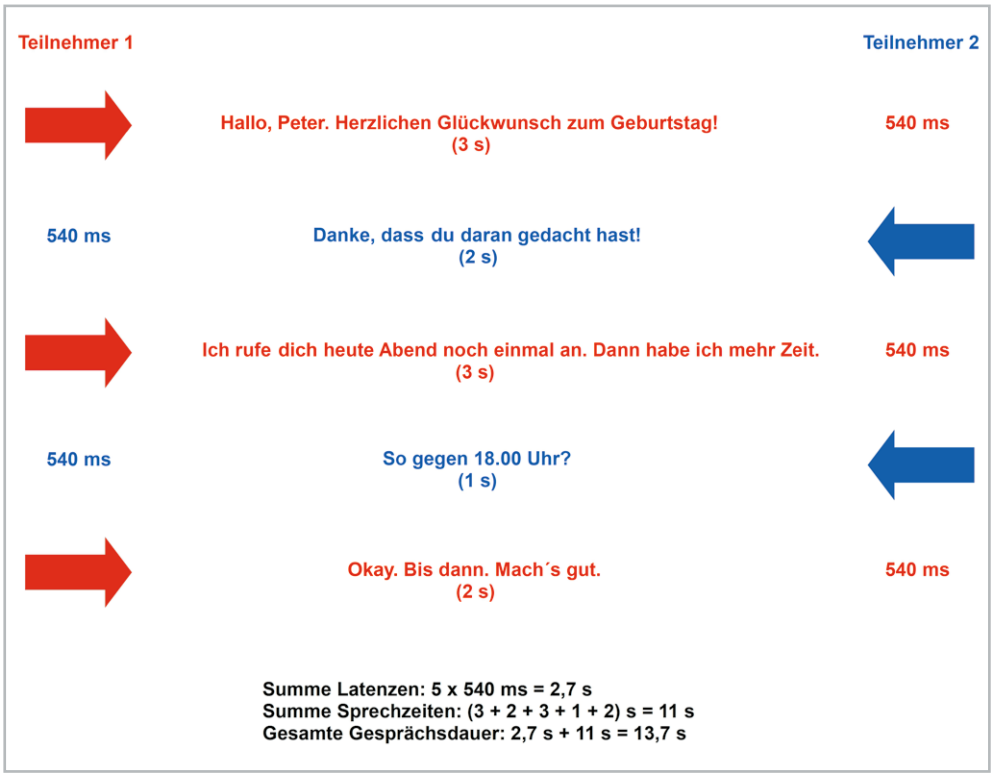

Bild 4: Latenzen vermindern die Übertragungsgeschwindigkeit.

etwas langsam im Verständnis des ihm Mitgeteilten ist und dementsprechend lange braucht, um seine Antwort zu formulieren, erzeugt damit eine zusätzliche Latenz, die den zügigen Gesprächsablauf unnötig behindert und den Informationsaustausch verlangsamt. Und das insbesondere dann, wenn die Antwort mangels Verständlichkeit oder Klarheit Rückfragen und deren Beantwortung erfordert. Als die Telefonie noch in den Kinderschuhen steckte, wurde dafür der Begriff „eine lange Leitung haben" geprägt.

Latenzen in der Musik

Wenn eine Gruppe von Musikanten zusammen spielt und singt, liegen die größten räumlichen Abstände zwischen den Mitgliedern meist im Bereich weniger Meter. Bei einer Schallgeschwindigkeit von 340 m/s dauert es also maximal 10 bis 20 ms, bis das Gespielte bei allen Musikern angekommen ist. Damit ist eine zufriedenstellende Synchronisation aller Beteiligter möglich. Schwieriger wird es schon bei großen Chören und Orchestern. Hier verhindern die akustischen und visuellen Verhältnisse die Selbstkoordination der Musiker allein über Hör- und Sichtkontakte. Um die Präzision des Zusammenwirkens zu erhöhen, bedarf es eines Dirigenten. Neben diesem rein technischen Aspekt des zeitlichen Zusammenführens der Ensemblemitglieder bringt der Dirigent natürlich auch seine künstlerischen Interpretationsvorstellungen für alle verbindlich in das konzertante Miteinander ein.

In der Musikproduktion und beim anspruchsvollen Homerecording führt heute kein Weg mehr am Computer vorbei. Im Zusammenwirken mit geeigneter Hard- und Software wird er zur „Digital Audio Workstation" (DAW). Die bei der digitalen Verarbeitung der akustischen und elektronischen Eingangssignale auftretenden Latenzen sind hierbei als Zeiten zwischen der Erzeugung von Sounds und deren Erreichen unseres Ohrs zu verstehen. Ein einzelner akustischer Konzertgitarrist hat damit keine Probleme. Das Schallloch seiner Gitarre ist einen knappen halben Meter von seinen Ohren entfernt, was eine Latenzzeit bis zum Eintreffen des Schalls von weniger als 2 ms bewirkt. Da nach vielen psychoakustischen Untersuchungen Künstler mit gutem Gehör Latenzen ab 11 ms als störend zu empfinden beginnen, kann diese Situation bereits beim Zusammenspiel mit einem weiter als 3,5 m entfernten Begleitinstrument entstehen.

Nicht umsonst tragen die Mitglieder vieler Bands „In-Ohr-Kopfhörer" (Bild 5), die ihnen ein möglichst ungestörtes latenzfreies Mithören des eigenen Instruments und des Outputs der anderen ermöglichen (in-ear monitoring). Idealerweise erfolgt das individualisierte Zuspielen des Kopfhörersignals rein analog, um die mit der Digitalisierung verbundenen Latenzen zu vermeiden. Standard-Bluetooth-Technik erhöht die Latenz im Bereich von 100 ms.

Aber auch rein digitale DAWs auf PC-Basis lassen sich in Hinblick auf Latenzen enorm optimieren. Dazu gehört der Einsatz von mehr RAM-Speicher, Defragmentierung der Festplatte, getrennte Festplatten für Betriebssystem und Nutzdaten, optimierte Au-

Bild 5: Gutes In-Ear-Monitoring bringt den Instrumentenklang nebengeräuschfrei und unverzögert in das Ohr des Musikers. Foto: Thomas Kiessling

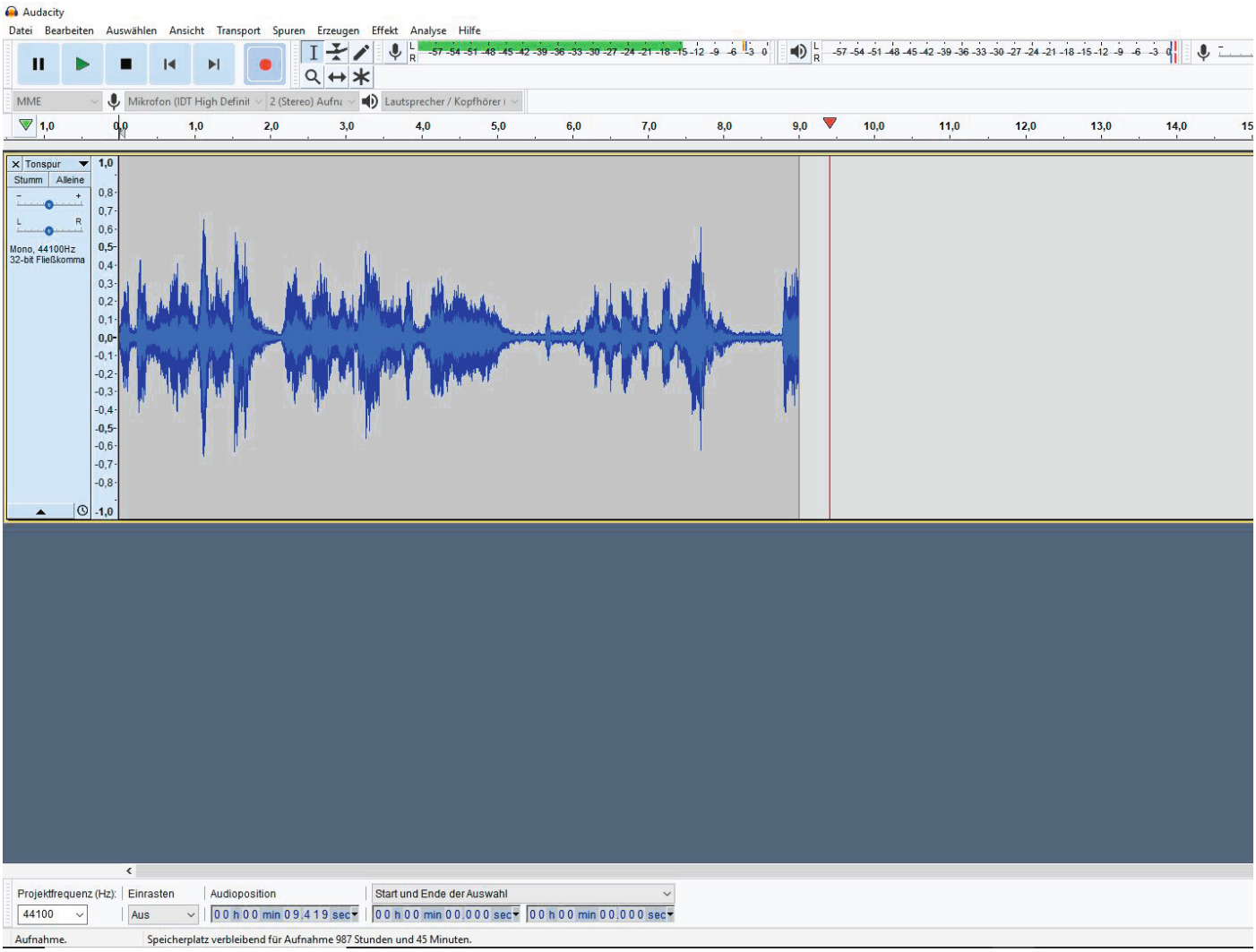

Bild 6: Einen guten Einstieg in die Technik der Digital Audio Workstations bietet das kostenlose Programm Audacity.

diotreiber, Aktualisierung der Firmware, Minimieren der Zahl der Hintergrundprozesse, Abschalten unbenötigter Spuren, Erhöhen der Puffergröße usw. Kann man die Latenzen mit diesen Maßnahmen nicht in den einstelligen ms-Bereich bringen, sind Investitionen in eine leistungsfähigere Hardware (Motherboard, Prozessor, Soundkarte, Digital Signal Processor ...) unausweichlich.

Bild 7: Die ersten Taschenuhren kamen wegen ihrer mäßigen Ganggenauigkeit mit einem Stundenzeiger aus. Quelle: Wikipedia

Einen guten Einstieg in das Thema DAW eröffnet die Freeware Audacity für PCs mit allen gängigen Betriebssystemen (Bild 6). Sie bietet ein virtuelles Tonstudio zur Bearbeitung von Audiomaterial aus zahlreichen Quellen (Mikrofon, Tonabnehmer, MIDI (Musical Instrument Digital Interface = digitale Musikinstrumentenschnittstelle), Festplatte) in vielfältigen Formaten.

Zeitsynchronisation – (auch) ein Latenzproblem

Der Zugang zu einer genauen Zeit war früher nicht einfach. Die Menschen hatten (besonders außerhalb ihrer Wohnung) meist keine Uhr und waren auf Zeitabschätzungen auf der Grundlage des Sonnenstandes angewiesen. Deshalb waren Kirchtürme regelmäßig mit Uhr und Glocken ausgestattet, welche die Menschen in einem gewissen Umkreis visuell und akustisch über den Stand der Zeit informierten. Die größere Reichweite hatte meist der schallgetragene Glockenschlag zur vollen, halben und viertel Stunde.

Schall hat eine Ausbreitungsgeschwindigkeit in der Umgebungsluft von 343 m/s. Er benötigt daher zum Überwinden einer Distanz von einem Kilometer knapp 3 s. Wenn unseren Vorfahren bei der Feldarbeit durch die Glocken der mehr oder weniger weit entfernten Dorfkirche das Ende des Arbeitstages eingeläutet wurde, spielten entfernungsabhängige zeitliche Wahrnehmungsdifferenzen des Geläutes kaum eine Rolle. Das Glockenläuten wurde als Referenz zum Stellen der Uhr verwendet.

Mit der verbleibenden Ungenauigkeit konnte man gut leben. Man kannte zwar das Phänomen des entfernten Blitzeinschlages, der augenscheinlich gleichzeitig, unabhängig von seiner Entfernung zum Einschlagort, sichtbar war, dessen donnernder Begleitschall aber erst Sekunden später am Ohr des Beobachters eintraf. Die unterschiedlichen Ausbreitungsgeschwindigkeiten von Licht und Schall waren somit bekannt, spielten im Alltagsleben aber keine Rolle.

Die älteste bekannte Abbildung einer Taschenuhr zeigt ein 1532 entstandenes, äußerst akkurates Gemälde Hans Holbeins des Jüngeren vom Danziger Hansekaufmann Georg Gisze in seinem Londoner Kontor (Bild 7). Als gangbestimmendes Element wurde bei diesem Uhrentyp eine sogenannte Unrasthemmung verwendet, die einen Fehlgang von nicht weniger als 15 bis 30 Minuten/Tag ermöglichte. Dieser Uhrentyp war deshalb nur mit einem Stundenzeiger ausgestattet.

Erst als dem Briten John Harrison um die Mitte des 18. Jahrhunderts der Bau von Uhren mit einer Ganggenauigkeit von wenigen Sekunden im Monat gelang, war das Zeitalter der Schiffschronometer angebrochen. Mit ihnen wurde die Navigation durch die Ableitung der geografischen Länge aus der Differenz zwischen Chronometerzeit und der durch Sternenpeilung ermittelten Ortszeit enorm vereinfacht und in der Genauigkeit gesteigert.

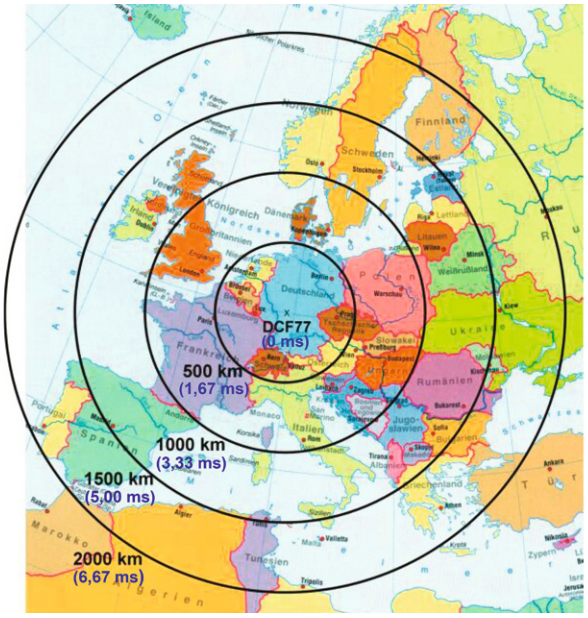

Bild 8: Die PTB Braunschweig verbreitet über den 75-kHz-Sender Mainflingen die amtliche Zeit der Bundesrepublik Deutschland in einem Umkreis von wenigstens 2000 km. Quelle: PTB

Rundfunk

Mit dem Aufkommen des Hörfunks wurde die exakte Zeiteinstellung von Taschen- und Wanduhren bis auf Sekundenbruchteile für jedermann möglich. Die jetzt gemachten Durchsagen (oft begleitet durch Zeitzeichen der Art „Mehrere kurze Töne, gefolgt von einem längeren", zur Kennzeichnung des besagten Punkts auf dem Zeitstrahl) erlaubten den Menschen überall im Sendebereich, ihre Uhren relativ exakt auf die gleiche Zeit zu stellen. Erst diese Synchronisation erlaubte Pünktlichkeit, also das exakte Einhalten vereinbarter Termine. Zu Zeiten des analogen Fernsehens erfüllte die allabendliche „Tagesschau-Uhr" die gleiche Funktion. Wenn ihr Sekundenzeiger in die Zwölf rutschte und der Gong ertönte, war es 20 Uhr. Das ist heute im Zeitalter der Digitaltechnik leider nicht mehr so, weder im DAB-Hörfunk noch im DVB-Fernsehen. Digitaltechnik und lange Satellitenverbindungsstrecken erzeugen Verzögerungen, die bis zu 10 s betragen können. In dieser Beziehung hat das auslaufende analoge „Echtzeit"-UKW-Radio einen klaren Vorteil.

Zeitzeichensender

Seit der Verbreitung elektronischer Uhren und Maschinensteuerungen erlauben Zeitzeichensender durch die Verbreitung eines maschinenauswertbaren Zeitzeichenfunksignals deren automatische Synchronisation mit einer Referenzuhr. Meistens werden „Funkuhren" einmal am Tag durch den Zeitzeichensender gestellt und laufen dann bis zur nächsten Synchronisation quarzgesteuert frei weiter. Dabei sind jederzeit Ganggenauigkeiten von wenigen ppm (ppm: part per million = 10^{-6}) garantiert.

Die Gründe für die Notwendigkeit einer synchronisierten, d. h. überall verfügbaren gleichen Zeit sind vielfältig. Man denke an Transaktionsprotokolle in verteilten Datenbanken, elektronischen Börsenhandel, kryptografisch zertifizierte sichere Dokumentenzeitstempel, Luftverkehrsüberwachung und Positionsmeldung, Multimedia-Synchronisation in Echtzeit-Telekonferenzen, Netzwerküberwachung, -messung und -steuerung, Edge-Computing usw. In zeitkritischen IoT-Anwendungen ist es unbedingt notwendig, einem Ereignis einen genauen Zeitpunkt zuordnen zu können.

DCF77

Der deutsche Zeitzeichensender DCF77 steht in Mainflingen im Kreis Offenbach, also in der Mitte Deutschlands (Bild 8). Er überträgt auf einer Trägerfrequenz von 77,5 kHz im Minutenrhythmus ein das Datum und die Uhrzeit enthaltendes Zeitzeichen. Dieses repräsentiert die gesetzliche Zeit Deutschlands und wird von einer Atomuhr der Physikalisch-Technischen-Bundesanstalt (PTB) in Braunschweig erzeugt. Die Bodenwelle des Senders ist jederzeit in einem Umkreis von 1000 km zu empfangen, die Raumwelle kann mit einmaliger Reflexion an der Ionosphäre in bis zu 2000 km Umkreis genutzt werden. In den Nachtstunden reicht sie wegen der multiplen Reflexionen an der Ionosphäre um ein Vielfaches weiter. Die Laufzeit des Zeitzeichensignals bis zum 2000-km-Umkreis des Senders beträgt 6,7 ms, d. h., Funkuhren in 2000 km Entfernung zum Sender gehen direkt nach der Synchronisation gegenüber denen in seiner unmittelbaren Nachbarschaft um 6,7 ms nach. Das lässt sich korrigieren, indem man die Entfernung zwischen Funkuhr und Zeitzeichensender über deren geografische Koordinaten ermittelt und den laufzeitbedingten verzögerten Eingang des Zeitzeichens entsprechend kompensiert. Bei den normalen Haushaltsfunkuhren wird das Zeitprotokoll nur mit einer Genauigkeit von typisch 10 bis 100 ms ausgewertet (niedrige Kurzzeitgenauigkeit!), sodass hier die entfernungsabhängige Laufzeit keine wesentliche Rolle spielt. Professionelle Funkuhren mit optimierten Dekodieralgorithmen und nicht zu schmalbandigen Empfangsfiltern stellen den exakten Beginn einer Sekundenmarke mit weniger als 100 µs genau fest. Will man von dieser Genauigkeit Gebrauch machen, ist die Berücksichtigung der entfernungsabhängigen Laufzeit des Empfangssignals sehr wohl erforderlich.

NTP

Schon zu Zeiten des Internetvorläufers ARPANET (Advanced Research Projects Agency Network) begann man sich Gedanken über die Synchronisation von Netzteilnehmern zu machen. Das ARPANET ging aus einer 1968 begonnenen Entwicklung des MIT (Massachusetts Institute of Technology) im Auftrag des amerikanischen Verteidigungsministeriums (DoD: Department of Defense) hervor und legte die technologischen Grundlagen des heutigen Internets. Dabei kam erstmals das Konzept der Paketvermittlung zum

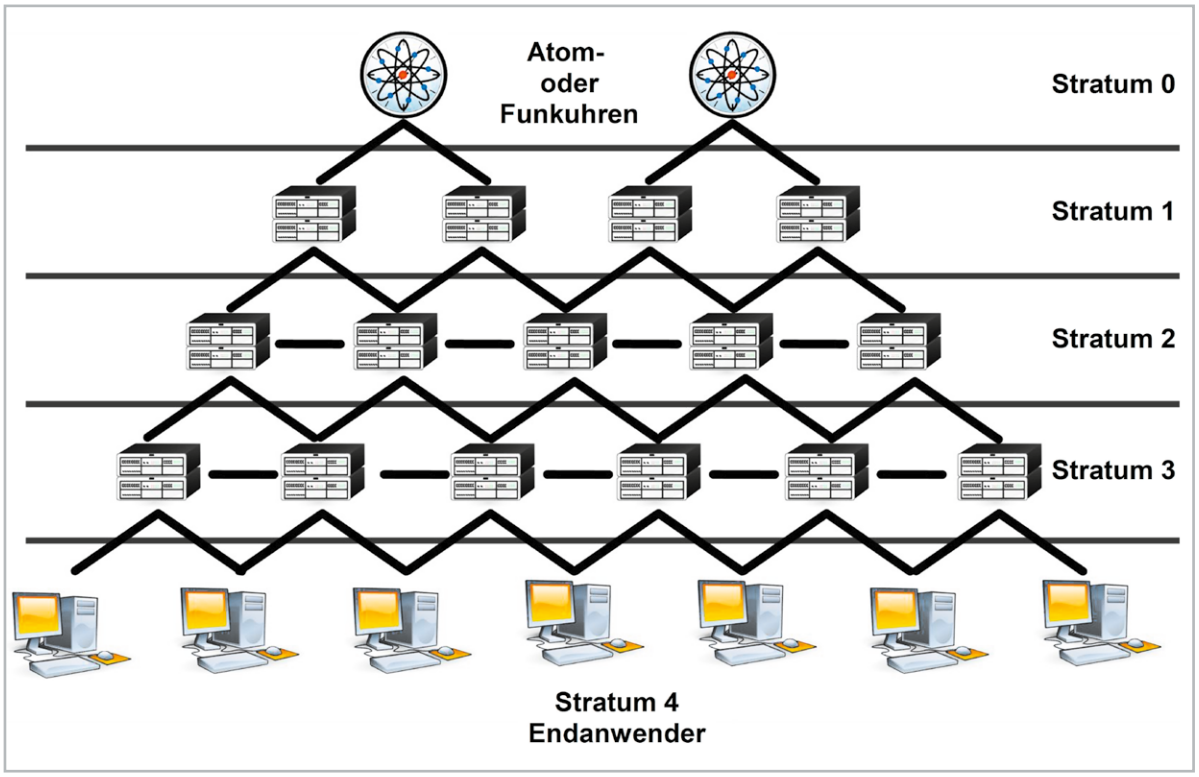

Bild 9: Das Network Time Protocol (NTP) dient in IP-Netzen der hochgenauen Zeitübermittlung von hierarchischen Zeitservern an nachfragende Klienten. Quelle: Wikipedia

Einsatz, bei dem die Kommunikationsinhalte nicht wie bisher üblich leitungsvermittelt, sondern in Pakete aufgeteilt über freie Netzwerkpfade von der Quelle zur Senke übertragen werden.

Das Network Time Protocol (NTP) wurde von David Mills an der University of Delaware entwickelt und geht in seinen Ursprüngen auf RFC 958 im Jahr 1985 zurück [1]. Heute, nach über dreißig Jahren wird NTP

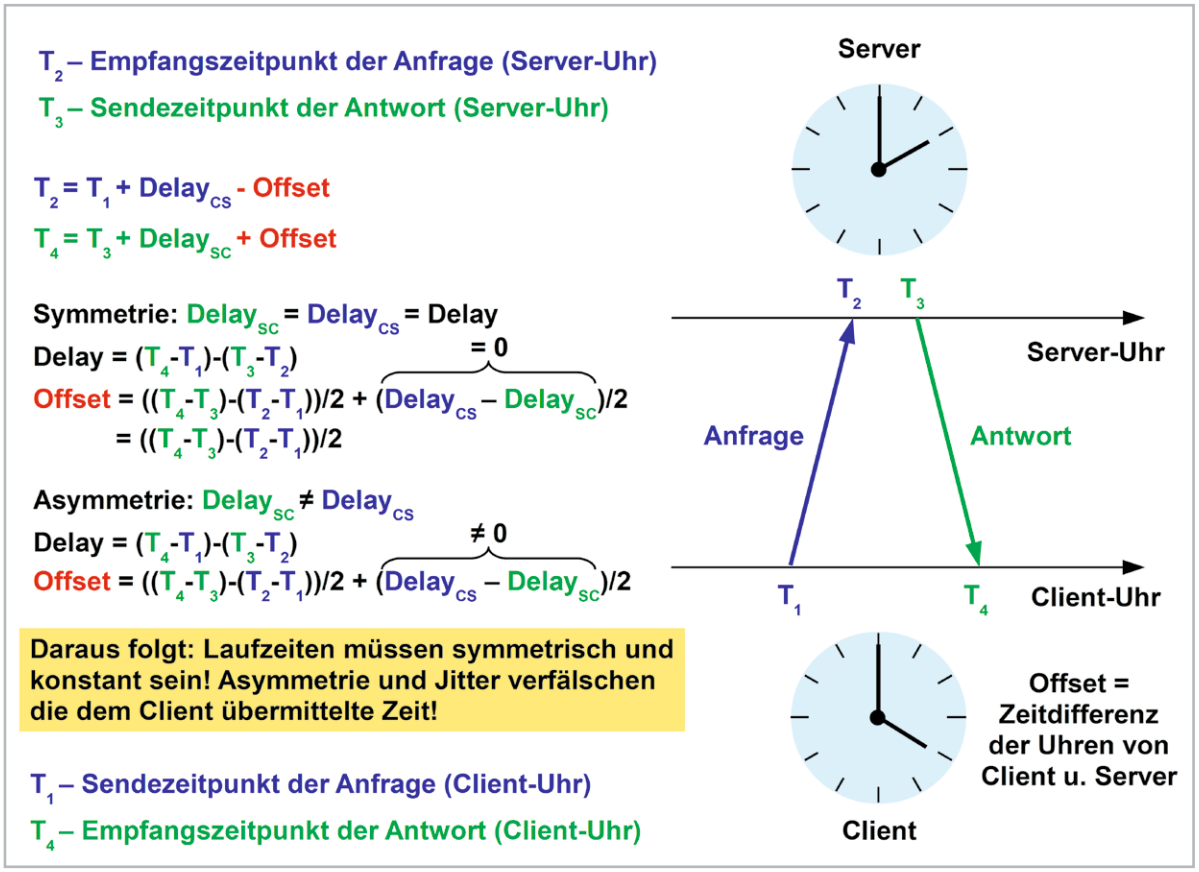

Bild 10: Über den Austausch von NTP-Paketen zwischen Klienten und Servern kann die Klientenuhr gestellt werden, indem sie die Serverzeit übernimmt. Die Paketlaufzeiten werden dabei herausgerechnet.

immer noch in der von RFC 5905 beschriebenen Version 4 aus dem Jahr 2010 verwendet. Ab 2010 berücksichtigt RFC 5908 die Erweiterung des Internetadressraums gemäß IPv6.

Die Zeitinformation bezieht ein Netzwerkrechner (NTP-Client) von einem NTP-Server über einen 64 Bit langen Zeitstempel. 32 Bit werden dabei für die Angabe der seit dem 1. Januar 1900, 00:00:00 Uhr vergangenen Sekunden verwendet, die verbleibenden 32 Bit geben den Sekundenbruchteil an. Somit lassen sich bis etwa zum Jahr 2036 ($1900 + 2^{36}$ s) Zeitstempel mit einer Auflösung von 0,23 ns (2^{-36} s) generieren. Den Zeitüberlauf bei 136 Jahren (era = Ära) muss das Betriebssystem eines Rechners berücksichtigen.

Die NTP-Topologie besteht aus einer Hierarchie von primären Servern (Stratum 1, Stratum = Schicht), die ihre Zeitinformation direkt von einer oder mehreren hochgenauen Zeitquellen (Stratum 0: Funkuhr, Atomuhr, GPS ...) beziehen und diese an hierarchisch gestaffelte sekundäre Zeitserver (Stratum 2, Stratum 3 ...) weitergeben (Bild 9). Die NTP-Server der verschiedenen Schichten kommunizieren miteinander und steigern dadurch die Genauigkeit ihrer lokalen Zeiten. Am Ende der Hierarchie stehen LAN-NTP-Server, deren Zeitinformation von den LAN-Clients der Teilnehmer abgefragt wird. So wird der Datenverkehr auf den Netzen reduziert und trotzdem eine hochwertige Verbreitung einer möglichst genauen Uhrzeit im LAN ermöglicht.

Das Prinzip der Zeitübermittlung zwischen Server und Client kann in Bild 10 nachvollzogen werden. Es wird angenommen, dass sowohl Server als auch Client über eine Uhr verfügen. Zwischen beiden Uhren besteht eine Zeitdifferenz (Offset). Der Client möchte nun die genauere Zeit des Servers übernehmen und fragt deshalb mit einem Datenpaket, welches seinen Absendezeitpunkt Client-Uhr (T1: Originate Timestamp) enthält, beim Server an. Der Server wertet das Paket aus und schickt es, um die Zeiten von Paketein- und -ausgang gemäß Server-Uhr (T2: Receive Timestamp, T3: Transmit Timestamp) ergänzt, an den Client zurück. Dieser registriert den Paketeingang gemäß seiner Uhr (T4) und ist nun im Besitz der vier Zeiten T1, T2, T3 und T4. Unter der Voraussetzung, dass die Paketlaufzeiten (Delay) vom Client zum Server (DelayCS) und vom Server zum Client (DelaySC) gleich sind (Symmetrie), spielen sie bei der Bestimmung des Offsets keine Rolle. Durch Berücksichtigung des Offsets vollzieht die Client-Uhr die Übernahme der Serverzeit. Richtungsabhängig unterschiedliche Paketlaufzeiten (Asymmetrie) und Schwankungen der Paketlaufzeiten (PDV: Packet Delay Variation = Jitter) verfälschen hingegen die vom Server übermittelte Zeit. Die Genauigkeit der NTP-Synchronisierung hängt von vielen Einflussfaktoren ab, insbesondere der Hardware, dem Betriebssystem und der Qualität der Netzwerkverbindung. Sie liegt im Millisekundenbereich.

Eine signifikante Genauigkeitssteigerung bis in den Submikrosekundenbereich ist mit dem im Jahr 2002 als IEEE-1588-Standard definierten „Precision Clock Synchronisation Protocol for Networked Measurement and Control Systems", oder kurz Precision Time Protocol (PTP), möglich. Dabei wird über einen Algorithmus der Zeitserver ermittelt und als Referenz verwendet, der die exakteste Zeit ausgibt (Best Master Clock). Die Anforderungen an das Netzwerk und die Rechnerressourcen für die lokale Uhr sind dabei relativ gering. Die letzte Weiterentwicklung geht als PTPv2 (IEEE 1588-2008) auf das Jahr 2008 zurück.

GPS

Millionenfach wird heute mit Empfängern für den Empfang der Sendesignale der erdumkreisenden Flotte von mindestens 24 aktiven GPS-Satelliten in 20.000 km Höhe navigiert (GPS: Global Positioning System = weltweites Positionsbestimmungssystem). In deren Sendesignalen sind hochgenau die jeweiligen Positionsdaten und die GPS-Systemzeit zum Sendezeitpunkt codiert. Aus den Signallaufzeiten von drei Satelliten ist die Position des Empfängers prinzipiell bestimmbar.

Weil der Empfänger aber in der Praxis nicht über eine eingebaute Referenzuhr mit der erforderlichen Genauigkeit verfügt, benötigt er zur exakten Laufzeitermittlung das Signal eines vierten Satelliten. Jetzt kann er seine eigene Position und Geschwindigkeit zum Empfangszeitpunkt berechnen. Tatsächlich empfängt ein GPS-Receiver zwischen sechs und zwölf Satelliten gleichzeitig (abhängig von der Konstellation der Satelliten am Empfangsort) und steigert damit die Präzision der Laufzeitbestimmungen auf Nanosekunden und der daraus abgeleiteten Ortungsgenauigkeit auf wenige Meter genau (Bild 11).

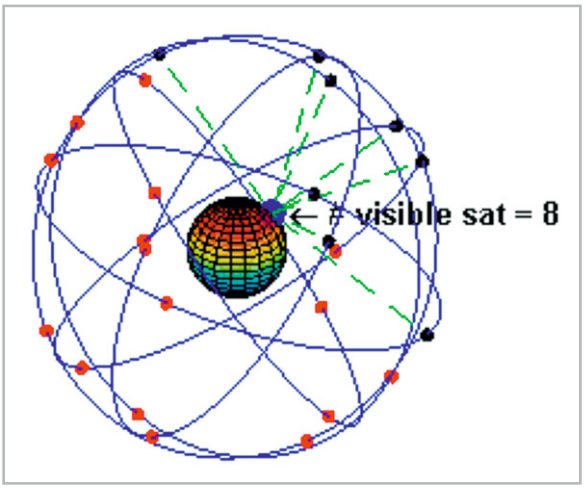

Bild 11: Hochgenaue Atomuhren an Bord der GPS-Satelliten sind die Grundlage weltumspannender Navigation und versorgen den GPS-Receiver zugleich mit einer hochgenauen Zeit. Quelle: Wikipedia

Da jeder GPS-Empfänger also fortlaufend über exakte Zeitinformationen verfügt, kann er problemlos zur Synchronisation von Uhren, Steuerungen und anderen verteilten technischen Systemen herangezogen werden. Die Genauigkeit der ermittelten Zeit ist unabhängig von der Position von GPS-Empfänger und Uhr/Steuerung auf der Erdoberfläche und liegt je nach getriebenem Aufwand bei ± 1 µs bis zu ± 10 ns (hohe Kurzzeitgenauigkeit!).

Zusammenfassung

Es wurde dargestellt, dass Signallaufzeit, Reaktionszeit und Verarbeitungszeit Latenzen verursachen, welche die Kommunikationsgeschwindigkeit einschränken und die Herstellung von Synchronität erschweren. In Teil 2 dieser Folge geht es konkret um das Bluetooth-Verfahren zur drahtlosen Datenübertragung über kurze Distanzen. Das Verfahren wurde in den 1990er-Jahren als Industriestandard IEEE 802.15.1 von der Bluetooth Special Interest Group entwickelt und setzte sich schnell für die einfache funkbasierte Audioübertragung durch. Hi-Fi-Fans rümpften aber zunächst die Nase, denn die eingesetzten Kompressionsverfahren schlugen auf die Klangqualität durch. Die Bluetooth-Weiterentwicklung durch den Halbleiterhersteller Qualcomm zum aptX-Standard schafft seit 2016 Abhilfe. ELV

Weitere Infos:

[1] https://tools.ietf.org/html/rfc958