Stephan Schultze: “Fehlertolerantes Kommunikationssystem für hochdynamische Antriebsregelungen.“ Darmstädter Dissertation 1995 (Übersicht)

Heute stehen für fast alle Anwendungsgebiete jeweils optimierte **Einzel**antriebe mit **digitaler** Signalverarbeitung (µ-Prozessoren, ASICs) zur Verfügung. Die Funktionen, die auf dieser Einzel-Leitebene realisiert werden, haben sich stetig in Richtung Diagnose, Identifikation, Selbsteinstellung usw. erweitert ("intelligenter Antrieb"). In zahlreichen technischen Prozessen (z.B. Bearbeitungsvorgänge mittels Werkzeugmaschinen) müssen **mehrere** Antriebe koordinierte Bewegungen ausführen. Die Koordinierungsfunktion obliegt in der Regel einer hierarchisch überlagerten Gruppen-Leitebene ("Numerische Steuerung"), die bekanntlich ebenfalls digitale Signalverarbeitung verwendet.
Beide angesprochenen Leitebenen haben heute einen beachtlichen Entwicklungsstand erreicht; der Informationsaustausch (Kommunikation) zwischen diesen Leitebenen ist jedoch erst seit vergleichsweiser kurzer Zeit zum Gegenstand von Forschung, Entwicklung und Standardisierung geworden. Gegenüber dem immer noch oft verwendeten Informationsaustausch über analoge (±10V) und binäre Signale auf jeweils einzelnen Leitern, bietet die digitale Kommunikation erhebliche Vorteile hinsichtlich Flexibilität (z.B. bei Parametersätzen und Diagnosedaten), hohe Genauigkeit und Reproduzierbarkeit sowie reduziertem Verkabelungsaufwand. Erste digitale Kommunikationssysteme in diesem Bereich befinden sich in der Markteinführungsphase bzw. am Beginn der Marktdurchdringung. Im störungsfreien Betrieb bewältigen sie die gestellten Aufgaben zufriedenstellend; ihre Fähigkeit auch beim Vorhandensein von Fehlern die die geforderte Funktion des Gesamtsystems (z.B. Herstellen eines Werkstückes) aufrecht zu erhalten, ist nicht besonders hoch entwickelt. Da aber die Ansprüche bezüglich Zuverlässigkeit und Verfügbarkeit wachsen werden, befaßt sich Herr Schultze in der vorgelegten Arbeit mit der Aspekt der Fehlertoleranz in obigem Anwendungsgebiet.
In den einführenden Kapiteln stellt Herr Schultze die zeitlichen Anforderungen an die Kommunikation bei räumlich verteilten Antriebsregelungen (z. B. in Werkzeugmaschinen) zusammen. Dabei sollen möglichst keine zeitlich variable, d. h. statistisch schwankende Totzeiten in die Regelkreise der Antriebsregelung eingefügt werden. Die kommunikationsbedingten Totzeiten sollen möglichst klein und konstant sein. Einige der heute am Markt existierenden Kommunikationssysteme für verteilte Antriebsregelungen werden beschrieben und dahingehend analysiert, wie weit sie die zuvor gestellten Forderungen erfüllen. Im einzelnen werden die in Deutschland für diesen Zweck vornehmlich benutzten Kommunikationssysteme „Profibus“, „Interbus S“ und „Sercos interface“ behandelt, wobei sich zeigt, daß die durchgehende Synchronisation zwischen allen Datenverarbeitungsfunktionen (z.B. Regler, Steuerungen...) und allen Kommunikationsfunktionen, so wie sie bei „Sercos interface“ verwendet wird, die zeitlichen Forderungen am besten erfüllt.
Nachdem zuvor im wesentlichen das zeitliche Verhalten der Kommunikationseinrichtung im störungsfreien Betrieb im Vordergrund stand, wird ab dem dritten Kapitel der zentrale Aspekt der Arbeit eingeführt, nämlich die Fehlertoleranz, also die Fähigkeit des Kommunikationssystems trotz Vorhandensein einer begrenzten Anzahl von Fehlern die geforderte Funktion des Gesamtsystems (z. B. Herstellen eines Werkstückes) aufrecht zu erhalten.
Um existierende Kommunikationssysteme einerseits, sowie das von Herrn Schultze in späteren Abschnitten vorgeschlagene neue System andererseits, hinsichtlich ihrem Verhalten bei Vorhandensein von Fehlern beurteilen zu können, wird mit Hilfe der sogenannten „Fehlervorgabe“ festgelegt, welche Arten von Fehlern ohne Beeinträchtigung des Prozesses (z. B. Herstellen eines Werkstückes) überstanden werden müssen. Neben den Einzel-Bit-Fehlern sowie Büschelfehlern sind dies insbesondere auch der Ausfall ganzer Übertragungsstrecken (z. B. physikalisches Durchtrennen eines Übertragungsmediums) sowie auch Fehler in den Teilnehmeranschaltungen dergestalt, daß diese unkontrolliert, d. h. ohne sich an das vereinbarte Protokoll zu halten, Informationen bzw. Störungen absenden.
Hinsichtlich ihres Verhaltens beim Auftreten der in obiger Fehlervorgabe spezifizierten Fehlerfälle werden im folgenden zwei existierende Bussysteme nämlich das „FDDI“ und das „Duplex Ringleitungssystem“ analysiert. Beide Bussysteme haben redundante Verbindungsstrukturen in Form von zwei Ringleitungen, wobei FDDI optische Übertragungsstrecken verwendet. Das Ergebnis dieser Analyse in Kapitel 4 zeigt, daß beide Systeme die vorgegebenen Fehler verkraften, also die für die gegebene Anwendung geforderte Fehlertoleranz aufweisen.
Das Problem bei diesen fehlertoleranten Systemen besteht jedoch darin, daß in vielen der Fehlerfälle das Echtzeitverhalten nicht ausreichend ist, d. h. beim Auftreten eines Fehlers, der die Umschaltung auf eine redundante Übertragungsstrecke erfordert, geht zu viel Zeit verloren, so daß der Prozeß (z. B. Fertigung eines Werkstückes) nicht störungsfrei fortgesetzt wird.

Nachdem in einem vorangegangenen Abschnitt das Echtzeitverhalten der Systeme „Profibus“, „Interbus S“ und „Sercos interface“ dargestellt wurde, werden diese Systeme jetzt an obiger „Fehlervorgabe“ gemessen. Dabei zeigt sich, daß keines der echtzeitfähigen Systeme die erforderliche Fehlertoleranz hat.
Somit liegt folgende Situation vor: Die existierenden fehlertoleranten Systeme haben ein unzureichendes Echtzeitverhalten (insbesondere beim Umschalten auf redundante Strukturen) und die existierenden Systeme, bei denen die Echtzeitforderungen erfüllt sind, haben unzureichende Fehlertoleranz.
Ziel des eigenen Vorschlags von Herrn Schultze, der im fünften Kapitel ausgearbeitet wird, ist daher ein echtzeitfähiges und gleichzeitig fehlertolerantes Kommunikationssystem.
Da „Sercos-interface“ im störungsfreien Fall für den vorliegenden Anwendungsbereich eine gut geeignete Kommunikationseinrichtung darstellt, werden alle seine positiven Eigenschaften übernommen. Im störungsfreien Fall verhält sich das neue System im Prinzip wie „Sercos-interface“, wobei es jedoch zusätzlich erweiterte Diagnose- und Konfigurationsmöglichkeiten zur Verfügung stellt. Um die geforderte Fehlertoleranz zu erreichen werden (wie bei FDDI) zwei gegenläufige Lichtwellenleiterringe (anstatt nur einem bei „Sercos-interface“) verwendet. Ziel ist es, daß beim Ausfall ganzer Übertragungsabschnitte (Streckenfehler) Rekonfigurationsmaßnahmen „in Echtzeit“ stattfinden, d. h. der Prozeß soll hiervon nicht beeinträchtigt werden. Als weitere, neue Eigenschaft wird angestrebt, daß einzelne Teilnehmer „Online“ mit in das Kommunikationssystem aufgenommen werden können. Das bedeutet, daß ohne Beeinträchtigung des Echtzeitverhaltens bereits arbeitender Antriebe eine strukturelle Erweiterung z. B durch Ein/Ausgabeeinheiten ermöglicht werden soll.
Die Maßnahmen zur Erreichung obiger Ziele werden im folgenden skizziert.

1.) Rekonfiguration bei Streckenfehler:

Bereits die Festlegung bei „Sercos-interface“ beinhalten, daß auf dem LWL-Ring spätestens nach 12 Bitzeiten eine Signalflanke kommen muß. Diese Flanken sorgen dafür, daß Phasenregelkreise zur Erzeugung des Empfangstaktes stets eingerastet bleiben. Bei dem neuen System dienen diese Flanken zusätzlich dazu, einen Bruch (Durchtrennung) des Ringes festzustellen. Alle dem Fehlerort nachfolgenden Teilnehmer erkennen nahezu gleichzeitig, daß die Flanken ausbleiben. Daraufhin erzeugt jeder Teilnehmer selbst an seinem Ausgang Signalflanken. Nur der Teilnehmer, an dessen Eingang auch danach keine Flanken ankommen, grenzt unmittelbar an die Fehlerstelle an. Dieser Teilnehmer weiß also, daß vor dem betreffenden Eingang der zugehörige Ring unterbrochen ist. Bei einem Einfachfehler erhält dieser Teilnehmer jedoch über den am anderen Ring angeschlossenen Eingang noch korrekte Information. Er bildet jetzt dadurch einen neuen Ring (Rekonfiguration) in dem er die vom ungestörten Eingang stammende Information auf dem Ausgang des gestörten Ringes weitergibt. Dieses Verfahren der Rekonfiguration basierend allein auf lokal vorhandene Informationen führt auch dann noch zum Erfolg, wenn im gleichen Übertragungsabschnitt beide Lichtwellenleiter unterbrochen werden. Dann bilden sich nach dem beschriebenen Verfahren automatisch zwei unabhängige Ringe. Die Grenzen der Redundanz sind jedoch dann erreicht, wenn beide Ringe so unterbrochen sind, daß die Unterbrechungsstellen in verschiedenen Übertragungssegmenten liegen. Dann erhalten die zwischen diesen Unterbrechungsstellen liegenden Teilnehmer auf keinem der Ringe irgendeine Eingangsinformation. Diese Teilnehmer können nicht mehr an der Kommunikation teilnehmen. Gleichwohl rekonfiguriert sich das restliche System, so daß wiederum zwei unabhängige Ringe mit den verbliebenen Teilnehmern gebildet werden.

2) Verwendung redundant empfangener Telegramme.

Auf beiden Ringen zirkulieren die gleichen Nachrichten, jedoch in entgegengesetzter Richtung. Aufgrund der physikalisch bedingten Laufzeit (Aufbereitung der Signale) laufen im fehlerfreien Zustand bei einem Antrieb die gleichen Informationen auf Ring 1 zu einem anderen Zeitpunkt als auf Ring 2 ein. Damit stellt sich die Frage, wann ein Telegramm als gültig angesehen wird. Herr Schultze untersucht die drei Möglichkeiten.

a) Verwendung des ersten korrekten Telegramms. Wird ein Telegramm mit fehlerfreier Prüfsumme und richtiger Länge empfangen so wird dieses gültig.

b) Verwendung zweier gleicher Telegramme. Nur wenn auf beiden Ringen zwei gleiche, korrekte Telegramme empfangen werden, wird dieses gültig. Dies führt zu einer sehr kleinen Restfehlerwahrscheinlichkeit, jedoch zu einer geringen Verfügbarkeit.

c) Verwendung nicht verschiedener Telegramme. Wird ein korrektes Telegramm auf einem Ring empfangen, so wird dies nur dann gültig, wenn auf dem zweiten Ring innerhalb einer durch die maximale Signallaufzeit im Ring kein weiteres korrektes Telegramm mit anderem Dateninhalt empfangen wird. Es wird im letzten Kapitel gezeigt, daß diese Methode ein Optimum bezüglich Sicherheit, Verfügbarkeit und Ausfallrate der Telegramme darstellt, so daß diese Methode im weiteren verwendet wird.

3.) Bedingungen zum Senden auf den Ringen

„Sercos-interface“ verwendet ein zeitschlitzgesteuertes Zugriffsverfahren d. h. jeder Teilnehmer wird während der Initialisierung darüber informiert zu welchem Zeitpunkt innerhalb des Zyklusses er

Informationen auf den Ring geben darf. Während der übrigen Zeit hört ein Teilnehmer die einlaufenden Informationen mit und gibt sie physikalisch aufbereitet an seinem Ausgang im Ring weiter. Der Zyklusbeginn, auf den sich alle Sendezeiten der einzelnen Teilnehmer beziehen, wird durch ein Synchronisationstelegramm des Masters (MST) den Teilnehmern mitgeteilt. Das neue System verwendet zwei gegenläufige Ringe, auf denen die Synchronisationstelegramme zu unterschiedlichen Zeitpunkten in einem Teilnehmer ankommen. Damit entsteht die Frage, zu welchem Zeitpunkt dieser Teilnehmer senden soll.
Bei der Beantwortung dieser Frage muß ein möglicher „Störer-Ausfall“ eines Teilnehmers berücksichtigt werden. Ein „Störer-Ausfall“ liegt vor, wenn ein Teilnehmer sich nicht an das vereinbarte Protokoll hält, sondern zu beliebigen Zeitpunkten beliebige Bitmuster sendet. Erschwerend wird hier angenommen, daß ein derartiger Ausfall auf beiden Sendeausgängen eines Teilnehmers vorliege. Eine Ausprägung dieses Zustandes wäre der Ausfall der Spannungsversorgung dieses Teilnehmers, dann wäre das am Ausgang gesendete Signal ständig Null (kein Licht). Dieser Fall ist aber trivial, da er bereits durch die zuvor behandelte Flankenerkennung abgedeckt wird, wobei die übrigen (gesunden) Teilnehmer sich so rekonfigurieren, daß sie zwei Ringe bilden. Daher nimmt Herr Schultze an, daß auf beiden Ausgängen des Störers ständig Flankenwechsel generiert würden. Die Flankenerkennung kann den Störer nicht ausgrenzen, da er ja hinreichend viele Flanken liefert. In diesem Fall erhalten die nachfolgenden Teilnehmer nicht mehr das vom Master gesendete Synchronisationstelegramm, da es durch den davorliegenden Störer zerstört wird.
Herr Schultze untersucht mehrere Möglichkeiten den Zeitpunkt des Sendens eines Teilnehmers - unter Berücksichtigung des Ausfalls der Synchronisationstelegramme - festzulegen. Es zeigt sich, daß das gleichzeitige Senden auf beiden Ringen, das sich an dem ersten empfangenen Master Synchronisationstelegramm orientiert, ein einfacher und durchaus brauchbarer Weg ist. Dabei wird zwar etwas Zeit verschenkt, die detaillierten Untersuchungen führen aber mit Bild 5.20 zu dem Ergebnis, daß im Bereich der üblichen Zykluszeiten < 4 ms die Anzahl der möglichen Teilnehmer dadurch nicht wesentlich eingeschränkt wird.
Beim „Störer Ausfall“ erhalten z.B. die links vom Störer liegenden Teilnehmer die Telegramme des Masters auf Ring 1 und die rechts vom Störer liegenden Teilnehmer auf Ring 2. Umgekehrt erhält der Master die Telegramme der links vom Störer liegenden Teilnehmer auf Ring 2 und die der rechts vom Störer liegenden Teilnehmer auf Ring 1. Damit ist nach wie vor der vollständige Informationsaustausch gewährleistet und der Master kann zusätzlich den Störer identifizieren.

Verfügbarkeit und Restfehlerwahrscheinlichkeit

In Kap. 5.3 wird die Wahrscheinlichkeit des Kommunikationsausfalles sowie die Restfehlerwahrscheinlichkeit rechnerisch ermittelt und deren Verlauf in Abhängigkeit der Bitfehlerwahrscheinlichkeit und Telegrammlänge angegeben für das System mit nur einem Ring („Sercos-interface“) und verschiedene Varianten der Doppelringsystems. Unterstellt man eine sehr geringe Bitfehlerwahrscheinlichkeit von $p_{BE}=10^{-9}$, dann ist auch bei nur einem Ring die mittlere zu erwartende Zeit zwischen Kommunikationsausfällen $MFBF_{KE}$ unproblematisch. Wird jedoch eine Bitfehlerwahrscheinlichkeit von $p_{BE}=10^{-6}$ angenommen, dann ist bei 1000bit / Zyklus im Einzelring im Mittel alle 16 Minuten mit einem Kommunilationsausfall ($\hat{=}$ Produktionsstillstand) zu rechnen, was absolut untragbar ist. Beim vorgeschlagenen Doppelringsystem ist ein derartiges Ereignis unter den gleichen Bedingungen im Mittel nur alle 32 Jahre zu erwarten.
Für Bitfehlerwahrscheinlichkeiten kleiner als $0{,}5*10^{-6}$ wird beim vorgeschlagenen System die Restfehlerwahrscheinlichkeit klein genug um die höchste Datenintegritätsklasse (I3) der DIN19244, Teil10 zu erreichen, während die Datenintegrität beim Einzelring immer geringer als Klasse I3 ist.

Versuchsaufbau

Herr Schultze hat das vorgeschlagene System als Laboraufbau in Hard- und Software realisiert, was ab Kap. 5.2.9 knapp beschrieben wird. Die Hardware basiert auf PC-Einsteckkarten, die neben einem eigenen Rechner und einem Kommunikationscontroller zusätzliche Hardware zum Erkennen der Synchronisationstelegramme sowie Zeitwerke zum Bit-genauen Senden enthalten. Vier derartige Rechnerkarten, die in zwei PC’s eingesteckt werden, bilden die Hardwarebasis. Per Software werden die Funktionen der Karten so festgelegt, daß ein Master und drei Antriebsanschaltungen zur Verfügung stehen. Diese Zahl ist mindestens erforderlich um die entworfenen Rekonfigurationsmaßnahmen tatsächlich ablaufen zu lassen.
Die umfangreiche Software läßt sich gliedern in Programme, die auf dem PC und solche die auf den Einsteckkarten ablaufen, wobei jeweils zwischen den Funktionen eines Masters und denen eines Antriebs zu unterscheiden ist. Auf dem PC läuft grundsätzlich eine Bedieneroberfläche, mit deren Hilfe sowohl Aktionen in den Antriebs ausgelöst werden können (Erzwingen bestimmter Fehlerzustände) als auch die Reaktionen und die Fehlerdiagnose des Systems dargestellt wird. Die Programme in den Einsteckkarten haben die volle Funktionalität von „Sercos-interface“ zu realisieren und zusätzlich die neuen Funktionen, insbesondere bei der Aufnahme eines neuen Teilnehmers in den Doppelring während des Betriebes der übrigen.

Mit Hilfe des hier nur grob beschriebenen Versuchsaufbaues wird das Verhalten des neuen Systems in Fehlerfällen demonstriert. Unterbricht man beispielsweise beide Lichtwellenleiter zwischen dem Antrieb 1 und 2 dann läuft die Kommunikation mit allen Teilnehmern trotzdem in Echtzeit weiter. Lediglich im ersten Kommunikationszyklus nach dem Streckenfehler können gestörte Telegramme enthalten sein. Fällt ein Telegramm in einem Zyklus aus, dann wird dies wie bei „Sercos-interface“ toleriert, der betroffene Antrieb verwendet für diesen einen Zyklus die alten Sollwerte weiter. Zusätzlich meldet Antrieb 1 bzw. 2 dem Master, daß er am Ring 1 bzw. 2 keine Signalflanken mehr erhalten hat. Die Diagnosefunktion des Masters erzeugt in diesem angenommen Fehlerfall die Meldung „Ring 1 nicht geschlossen. Ring 2 nicht geschlossen. Fehler zwischen den Teilnehmern 1 und 2.“. Entsprechende Reaktionen und Meldungen werden natürlich auch beim einfachen Streckenfehler und beim Teilnehmerausfall (z.B. „Störer- Ausfall“) erzeugt.

P. Mutschler