

**ANALYSE**



# **CASCADING DATA CORRUPTION**

/ Februar 2017

KASKADIERENDE FEHLER IN VERTEILTEN  
UND HETEROGENEN NETZWERKEN

---

# INHALT

---

1	EINFÜHRUNG UND RELEVANZ FÜR CPS	4
2	AKTUELLER STAND IN WISSENSCHAFT UND WIRTSCHAFT	6
3	FORSCHUNGSOPTIONEN	7
4	LITERATURVERZEICHNIS	9



EUROPÄISCHE UNION  
*Investition in unsere Zukunft*  
Europäischer Fonds  
für regionale Entwicklung



**EFRE.NRW**  
Investitionen in Wachstum  
und Beschäftigung

---

## Autoren

UNIVERSITÄT  
DUISBURG  
ESSEN

*Offen im Denken*

Stefan Gries  
Volker Gruhn  
Marc Hesenius  
Julius Ollesch

---

## Kontakt



Institut SIKoM+ | Bergische Universität Wuppertal

Rainer-Gruenter-Straße 21  
42119 Wuppertal

E-Mail: [kontakt@cps-hub-nrw.de](mailto:kontakt@cps-hub-nrw.de)

Telefon: +49 202 439-1026

Fax: +49 202 439-1037

[www.cps-hub-nrw.de](http://www.cps-hub-nrw.de) | [www.facebook.com/cpshubnrw](https://www.facebook.com/cpshubnrw) | [www.twitter.com/cpshub](https://www.twitter.com/cpshub)



EUROPÄISCHE UNION  
Investition in unsere Zukunft  
Europäischer Fonds  
für regionale Entwicklung



EFRE.NRW  
Investitionen in Wachstum  
und Beschäftigung

---

# 1 EINFÜHRUNG UND RELEVANZ FÜR CPS

---

Cyber-Physical Systems, kurz CPS, sind gekennzeichnet durch eine enge Verzahnung von physischen und informationstechnischen Prozessen [1]. Zentraler Bestandteil der Hardware eines CPS ist sowohl die Sensorik, die echtweltliche Prozesse überwacht und wahrnimmt, als auch die Aktorik, die wahrgenommene und andere Prozesse aktiv steuert und beeinflusst.

CPS sind in der Regel eine Zusammenstellung aus heterogenen und teilweise autonomen verteilten Systemen, die für einen gemeinsamen und/oder eigene Zwecke kooperieren. Die Netze sind dabei keiner starren Struktur unterworfen, sondern können sich zur Laufzeit verändern, indem einzelne Komponenten ausgetauscht, ergänzt, entfernt oder verändert werden [2]. Durch diese änderbare Struktur entsteht ein emergentes<sup>1</sup> Verhalten des Gesamtsystems [3]. CPS sind somit geprägt von Dynamik.

Somit handelt es sich bei CPS um verteilte, heterogene, teilautonome Systeme, die in einem Gesamtsystem zusammengefasst werden. Das Netzwerk, durch das die Systeme eine logische Einheit bilden, ist keiner vorgegebenen Struktur unterworfen, sondern lose gekoppelt und bietet keine starre Topologie.

CPS entstehen erst durch die Interaktion von Hardware und Software – keiner der Teile allein kann die Anforderungen an ein CPS erfüllen. Es entsteht ein Wechselspiel – die Hardware, beispielsweise Sensorik, versorgt vernetzte Systeme, betrieben durch Software, mit notwendigen Informationen zur Entscheidungsfindung. Die Kommunikation erstreckt sich hierbei über viele Knoten hinweg. Betroffene Entscheidungen spiegeln sich wiederum im Verhalten von Hardware wider, den so genannten Aktoren, die durch Software gesteuert werden. In diesem Wechselspiel tragen echtweltliche Prozesse zur Entscheidungsfindung in Software bei, deren Resultate wiederum echtweltlich sichtbar werden können.

Die Vielfalt der in CPS eingesetzten Sensorik und Aktorik nimmt immer weiter zu. Im Jahr 2010 waren durchschnittlich bereits 35% des Wertes eines Automobils solche elektrischen Komponenten [4], die zunehmend Aufgaben von hydraulischen oder mechanischen Bauteilen übernehmen. Insofern ist mit einer weiteren steigenden Verbreitung elektronisch gesteuerter Bauteile zu rechnen. Hierdurch nimmt auch die Heterogenität und Komplexität der Komponenten im Netzwerk und des Netzwerks selbst stetig zu.

CPS bestehen aus drei Typen von Knoten im Netzwerk: **Sensoren, Aktoren und Aggregatoren**.

- **Sensoren** zeichnen echtweltliche Daten auf und geben diese an andere Knoten weiter. Selbst treffen sie keine Entscheidungen, sondern dienen als „Sinnesorgane“ des Gesamtsystems.

---

<sup>1</sup> Als emergent bezeichnet man Systeme, deren Verhalten durch nicht weniger als die Betrachtung des Gesamtsystems, insbesondere nicht durch die Betrachtung eines Teilsystems, erklärt werden kann. Darum sind klassische Client-Server Systeme per Definition nicht emergent, da der Server die Hoheit über alle Prozesse hält. Das Verhalten des Gesamtsystems ergibt sich somit durch das Zusammenspiel der einzelnen Systemkomponenten und wird nicht durch eine zentrale Instanz bestimmt.

- **Aggregatoren** sind Komponenten im Netzwerk, meist [kleine] Server, die Sensorwerte erhalten, um diese zu verarbeiten, zu aggregieren und um Entscheidungen aus ihnen abzuleiten. Diese Entscheidungen werden durch Aktorik sichtbar.
- **Aktoren** erhalten Aktionsdaten von Aggregatoren und führen diese aus. Somit beeinflussen Aktoren wiederum echtweltliche Prozesse, die womöglich durch Sensorik wahrgenommen werden.

Durch die Weitergabe von Informationen im Netzwerk entstehen Abhängigkeiten zwischen den einzelnen Knoten. Entscheidungen eines Aggregators beruhen nicht mehr nur auf eigenen Daten, sondern auch auf Sensorwerten von entfernten Netzwerk-Knoten. Die durch den Aktor ausgeführte Aktion ist seinerseits also nicht nur von einem [oder mehreren] steuernden Knoten abhängig, sondern ebenfalls von deren Datenlieferanten. Es entsteht ein Abhängigkeitsbaum, der beliebig viele Ebenen enthalten kann. An Aktoren sichtbar werdende Entscheidungen können daher oft nicht nachvollziehbar wirken, da der Abhängigkeitsbaum in der Regel nicht bekannt nicht einfach konstruierbar oder ableitbar ist. Um Abhängigkeiten in Netzwerken im Bezug zur Semantik zu erfassen und nachzuverfolgen wurden bisher wenige Arbeiten veröffentlicht. Soweit uns bekannt, hat noch keine andere Forschungsgruppe zu dem Thema, wie semantische Abhängigkeiten zwischen ausgetauschten Nachrichten in verteilten Netzwerken erfasst werden können, veröffentlicht.

---

## 2 AKTUELLER STAND IN WISSENSCHAFT UND WIRTSCHAFT

---

Konventionelle Methoden zum Network Monitoring wie Cisco NetFlow [5], [6] oder cflow [7] sind ausschließlich in der Lage quantitative Eigenschaften oder Metadaten zu erfassen und auszuwerten. Diese Eigenschaften umfassen Informationen wie Traffic-Menge oder die Art der versandten Datenpakete. Sie sind nicht in der Lage semantische Abhängigkeiten zu erkennen. Andere Tools für Flow Monitoring sehen sich mit ähnlichen Problemen in dieser Domäne konfrontiert [8], [9]. In Kontrast zu diesen dargestellten Ansätzen versuchen Marashi et al. [10] Abhängigkeiten zwischen physischen Komponenten und Cyberkomponenten in CPS zu modellieren. Dieser Ansatz betrachtet nicht nur Netzwerk-Eigenschaften, sondern bezieht sich ebenfalls auf Inhalte. Durch die dynamische Struktur und die hier vor Laufzeit angewandte Modellierung von Abhängigkeiten wird der Ansatz der Emergenz in CPS allerdings nicht gerecht und erlaubt keine Identifizierung von Abhängigkeiten zur Systemlaufzeit.

---

## 3 FORSCHUNGSOPTIONEN

---

Knoten in verteilten Netzwerken arbeiten zusammen und tragen ihre Arbeitsleistung zu einem Gesamtergebnis bei. Insbesondere bei CPS interagieren Knoten und beeinflussen den jeweiligen Entscheidungsfindungsprozess durch die Zulieferung von Daten, beispielsweise in Form von Sensorwerten. Ohne einen korrekten Informationsfluss im Netzwerk werden zentrale Prozesse, die Aktionen bestimmen, gestört und das CPS kann seine Aufgaben nicht mehr erfüllen.

Hierdurch entsteht ein hoher Grad an Abhängigkeit zwischen den einzelnen Knoten. Insbesondere Aktoren sind von zuliefernden Knoten abhängig, die ihnen ein Verhalten vorgeben. Letztendlich handelt es sich bei Aktoren nur um ausführende Instanzen des Netzwerkes, die in der Regel selbst keine Entscheidungen treffen. Aggregatoren haben bereits zuvor Informationen zusammengetragen und verarbeitet, um das korrekte situationsabhängige Verhalten eines Aktors zu bestimmen. Betrachtet man diese Beziehung, wird klar, dass die Aktoren von der korrekten Funktionsweise der Aggregatoren abhängig sind, welche ihrerseits zusätzlich zu dieser Abhängigkeit weitere Beziehungen untereinander haben. Die Verarbeitung von Daten findet in CPS in der Regel nicht zentralisiert, sondern verteilt statt. Hierdurch tauschen auch Aggregatoren untereinander verarbeitete Daten aus, die Prozesse beeinflussen und Ergebnisse bestimmen. Die hierbei durchgeführte Datenverarbeitung wiederum ist ihrerseits in hohem Maße von Sensorik abhängig, die Informationen über die Umwelt erfasst und weitergibt.

**Insgesamt ergibt sich eine Abhängigkeit der Aktoren von den Aggregatoren, die ihrerseits sowohl untereinander, als auch von Sensorik abhängig sind.**

Bei CPS kann es sich sowohl um sehr kleine, lokal beschränkte, als auch global verteilte Netzwerke mit einer enormen Anzahl an Knoten handeln. Je größer diese Netzwerke werden, desto schwieriger lassen sich intuitiv Abhängigkeiten zwischen einzelnen Aktionen und Eingabedaten erkennen. Die genauen Abhängigkeiten entstehen durch das Zusammenspiel der einzelnen Knoten und können sich zur Laufzeit durch das emergente Verhalten von CPS verändern. Dem Architekten einzelner Bereiche eines CPS muss beim Entwurf nicht jede Abhängigkeit bekannt sein – es kann auch auf bereitgestellte und bereits aggregierte Daten anderer Knoten zugegriffen werden, um eigene Entscheidungen abzuleiten.

Insgesamt entsteht so zur Laufzeit des Systems eine unbekannte Abhängigkeitsstruktur zwischen Sensoren, Aggregatoren und Aktoren. Die Aufdeckung dieser Abhängigkeiten ist insbesondere für zwei Anwendungsfälle essentiell:

- **Ursache und Wirkung verstehen:**

CPS sind hochverteilte Systeme, die Nachrichten zwischen verschiedenen Knoten austauschen und verarbeiten. Bei der Verarbeitung von Nachrichten kann jeder Knoten unabhängig von anderen mit den Daten verfahren und somit Inhalte ergänzen, verarbeiten oder ergänzen. Die genauen Aktionen sind nur dem verarbeitenden Knoten selbst bekannt und können nicht ohne weiteres von anderen nachvollzogen werden.

In vielen Szenarien müssen allerdings Entscheidungen eines CPS nachvollziehbar und verifizierbar sein. Wenn eine Entscheidung die ein Knoten eines CPS oder das CPS im Zusammenspiel getroffen hat nicht mehr nachvollziehbar ist, können die Einflüsse auf die Entscheidung nicht mehr identifiziert werden. Dementsprechend ist es des Weiteren nicht möglich durch Beeinflussung der Ursachen Einfluss auf ein möglicherweise nicht erwünschtes Verhalten zu nehmen. Da in der Regel Entscheidungen durch das Zusammenspiel der einzelnen Netzwerkknotten getroffen werden, wird dieser Effekt

durch die Emergenz des Netzes verstärkt. Neu zum Netz hinzugefügte Knoten verändern den Entscheidungsfindungsprozess und erfordern eine neue Analyse.

- **Fehler zurückverfolgen:**

Da wie bereits beschrieben nicht immer klar ist, wie Entscheidungen und Aktionen, die aus diesen Entscheidungen resultieren, zustande kommen, ist es ebenfalls nicht möglich, im Fehlerfall die Ursachen für diesen Fehler schnell und effizient zu identifizieren. Kommt es beispielsweise aufgrund falscher Sensorwerte zu einer nicht angemessenen Aktion eines Aktors, so muss es möglich sein, den Zusammenhang zwischen Sensorwert und Aktion herzustellen. Durch die mehrfache Weitergabe und Aggregation von Sensorwerten und anderen Informationen wird die eigentliche Quelle von Daten verschleiert und ist beim ausführenden Aktor nicht sichtbar. Fehler werden jedoch oft erst beim Aktor sichtbar, da dieser eine realweltlich wahrnehmbare Aktion ausführt.

Um zu verstehen, warum ein Fehler aufgetreten ist oder warum eine Entscheidung getroffen wurde, benötigen wir mehr als ein Datenflussdiagramm. Datenflussdiagramme beschreiben nur den generell möglichen Datenfluss in einem System. Der genaue Datenfluss innerhalb eines bestimmten Szenarios eines Systems wird jedoch nicht dargestellt. Wir benötigen einen Abhängigkeitsbaum auf Informationsebene, der veranschaulicht, welche Daten für die Berechnung eines Ergebnisses verwendet wurde - nicht nur, welche Knoten dafür verwendet wurden. Da diese Abhängigkeiten zur Laufzeit des Systems erzeugt werden, ist es nicht möglich, sie vorher zu modellieren.

Wir müssen erfahren können, welche Knoten in einem speziellen Fall an dem Erstellungs-, Verarbeitungs- und Lieferprozess einer Information beteiligt waren. Ferner sollte es möglich sein, zu verfolgen, welche eingehenden Daten und Informationen diese Knoten verwenden, um ihre ausgehenden Informationen zu erzeugen. Dies ist eine große Herausforderung in einem heterogenen Netzwerk, das zur Laufzeit geändert werden kann. Ohne diese Information kann ein Fehler nicht oder nur sehr schwer zu seiner Quelle zurückverfolgt werden, um genau dort behoben zu werden.



---

## 4 LITERATURVERZEICHNIS

---

- [1] Edward A. Lee, „CPS Foundations“, New York, NY, 2010.
- [2] Q. Hangping, Y. Zhicai, Q. Jichuan, und L. Zhipeng, „Topological structure modeling and analysis for cyber-physical systems“, in *Information Technology and Artificial Intelligence Conference (ITAIC)*, 2014 IEEE 7th Joint International, 2014, S. 523–526.
- [3] J. S. Osmundson, T. V. Huynh, und G. O. Langford, „Emergent Behavior in Systems of Systems“, in *INCOSE International Symposium*, 2008, Bd. 18, S. 1557–1568.
- [4] K. Reif, *Bosch Autoelektrik und Autoelektronik - Bordnetze, Sensoren und elektronische Systeme*. 2011.
- [5] Cisco, „Introduction to Cisco IOS® NetFlow“. .
- [6] „Cisco IOS Flexible NetFlow Technology - Data Sheet“. .
- [7] S. Poznyakoff, „GNU cflow“, *GNU cflow*. [Online]. Verfügbar unter: <https://www.gnu.org/software/cflow/>. [Zugegriffen: 09-Jan-2017].
- [8] V. Sekar, M. K. Reiter, W. Willinger, H. Zhang, R. R. Kompella, und D. G. Andersen, „cSamp: A System for Network-Wide Flow Monitoring.“, in *NSDI*, 2008, Bd. 8, S. 233–246.
- [9] Z. Su, T. Wang, Y. Xia, und M. Hamdi, „CeMon: A cost-effective flow monitoring system in software defined networks“, *Comput. Netw.*, Bd. 92, S. 101–115, Dez. 2015.
- [10] K. Marashi, S. S. Sarvestani, und A. R. Hurson, „Quantification and Analysis of Interdependency in Cyber-Physical Systems“, in *2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshop (DSN-W)*, 2016, S. 149–154.

---

## Gefördert durch



**EFRE.NRW**  
Investitionen in Wachstum  
und Beschäftigung



**EUROPÄISCHE UNION**  
*Investition in unsere Zukunft*  
Europäischer Fonds  
für regionale Entwicklung