The Real world does not wait

Whether in industry, commerce, or everyday life: we are surrounded by real time applications. They have different standards regarding acceptable delay and respond with different levels of urgency when the actual delay is longer.

Photo: iStockphoto

COPY —— Heinz-Jürgen Köhler

Bump–bump, bump–bump, bu-bump-bump. A monitor shows a heart beating irregularly. A technological innovation opens up new possibilities in the analysis of cardiac arrhythmia: the real time MRI. Magnetic resonance imaging is a medical imaging procedure. Now, thanks to a development by Göttingen physician Jens Frahm, moving images from inside the body are possible in real time. Frahm's procedure, which won the European Inventor Award, uses special technology to achieve an imaging speed of up to 50 images per second: Only a portion of these 50 images are based on measurements; reconstructive algorithms compute the differences between the measured images and fill in the gaps between them.

From this type of high-tech application to a mass transit app for smartphones: real time applications are all around us. "Wherever computers interact with the real world today, 'real time' is typically behind it," stressed Björn Brandenburg of the Max Planck Institute for Software Systems. The head of the Real time Systems Group calls this a strong real-world impact. However, real time is defined differently in these various fields, and not keeping up with it has consequences of differing severity. Real time is commonly separated into hard real time and soft real time: with the former, exceeding the time limit has significant consequences; with the latter, only the quality of a service is reduced. In this distinction, Brandenburg finds at least the absoluteness of the assertion dubious: "There is harder real time and softer real time," said the software specialist. Whether the mass transit app shows that the trolley is five minutes late after a delay of 30 second or 60 seconds is hardly relevant. With the real time MRI, it is more a matter of the fluidity of the movement and not whether it is transmitted on a one-minute delay.

However, if Amazon reports that a delay of 100 milliseconds when loading the homepage costs one percentage point in revenue, then we are talking about a significant economic loss. And if an airbag in an accident or an airplane during an automated landing procedure responds to commands after a delay, the consequences are a matter of life and death. According to Brandenburg, the lines between hard and soft are blurred. "I always need to see what restrictions I can tolerate and what steps I may need to be ready to take to remove those restrictions."

INHERENT LATENCY

Real time systems are always subject to inherent latency. "This delay is an additive phenomenon," emphasized Brandenburg. Take the anti-lock braking system of a car: The sensors check traction, then transmit this information to the wheels, the actuators and, ultimately, to the car's electronic control unit, which may initiate countermeasures. All of these processes produce latencies that aggregate into an end-to-end delay. To minimize this, every step needs to be analyzed and optimized. Large processing capacity, such as that of quantum computers, is certainly helpful. Even edge computing, where certain data is processed at the edges of the system and does not need to be sent first to the core processor, can improve performance.

No less important is the path the data must travel. The distance and number of way-stations affect the transmission speed. Experts agree: Industry 4.0, automated driving systems and the Internet of Things, in which large volumes need to be sent over long distances, can only be implemented comprehensively with the powerful 5G mobile communications standard. Data will soon be sent along one particularly long path by EarthNow: The American startup founded in 2017 has announced plans for satellite-based real time earth observation. A comprehensive network of high-tech satellites and a powerful transmission system are designed to ensure that users can view events in real time, such as illegal fishing or emerging environmental disasters. Aerospace giant Airbus, which also wants to manufacture the satellites, and Bill Gates have invested in the company. However, the company has yet to announce when this service will be available.

PREDICTING WORST-CASE BEHAVIOR

At the Max Planck Institute for Software Systems in Kaiserslautern, Brandenburg is studying the worst-case behavior of real time systems. Is large processing capacity a guarantee of reliable real time behavior? "Yes and no," said Brandenburg. Large capacity certainly offers good performance on average, but it is not immune to downward shifts. "If a system has good averages, I often have a lot of very good and a few very bad figures," explained Brandenburg. These bad figures, however, cannot be allowed to occur when releasing an airbag. This is why Brandenburg is conducting what he calls worst-case execution time analyses. What are the worst possible figures from a system when executing a command? To find this out, he looks for the longest path at the programming level and tests it. "Generating a lot of load, finding bottlenecks and applying stress. It's a rather laborious process."

All this serves to optimize more critical real time systems such as airbags or aircraft control systems where delays have grave consequences. Because, as Brandenburg laconically puts it, "The real world doesn't wait for the computer to finish."

Brytes was founded by Hendryk Hosemann and four co-founders in Dortmund in 2017. The start-up is currently at the end of its pilot phase and plans to enter the market in 2019. Brytes currently has 11 employees, including software specialists, behavioral economists and data scientists.
www.brytes.de

“See the moments of decision”

Real time in e-commerce: How human understanding can be brought to online commerce as explained by Hendryk Hosemann, CEO of the startup Brytes.

What service does Brytes offer and what customers do you attract?
We offer online stores what is called psychographic online personalization. There are three steps: 1) We identify the visitor to an online store and track how they navigate the page; 2) We attempt to gain an understanding of the user: “What kind of person is this at this moment? What are they looking for? What do they want?” 3) We use this understanding and respond in real time by making use of, say, messages. Overall, it is a matter of better understanding the visitor in that specific situation.

We make websites more empathetic and bring human understanding to online communication. It is suitable for any company that places value on the customer experience. This definitely pertains not only to high-end consumer goods, but also mass markets. It also has a purpose in grocery shopping, since that is about habit and a feeling of familiarity.

What parameters do you use to analyze a user’s current mental state?
We use over 200 parameters to understand what drives the visitor’s decisions and purchases. These include scrolling and mouse movements – Are they moving in one direction? How fast are they? In what order are they looking at offers? Where are they pausing? Are they going from the overview page or from other recommendations? We talk of a decision journey, which is basically a broken-down customer journey with which you can see the moments of decision.

To what extent do you access other information the user leaves behind online?
Analyzing user behavior from earlier visits to our client’s online store or on the Internet in general is irrelevant to us. Our system reacts in real time to a visitor’s current behavior. Even if they were to come repeatedly or were to have just come from a competitor’s store, the system would not respond differently. And even if we could technically do that, we do not save the customer’s information.

How do you define real time and what do you do to accomplish it?
When we talk about real time, we are talking about milliseconds. Faster is not technically possible: We load as quickly as the page loads. Delay-free website performance is important to our customers and we adapt ourselves to that. Without a delay subjectively perceived by the user, we analyze their behavior and respond to it. This, of course, makes demands on infrastructure and server capacity. We also keep a content delivery network ready for our clients to help cushion peak loads.

Photo: Brytes, WZL/Krentz

PRODUCTION

“Prediction thanks to real time”

The current and soon-to-be-expected benefits of real time in production control as explained by Dr. Günther Schuh.

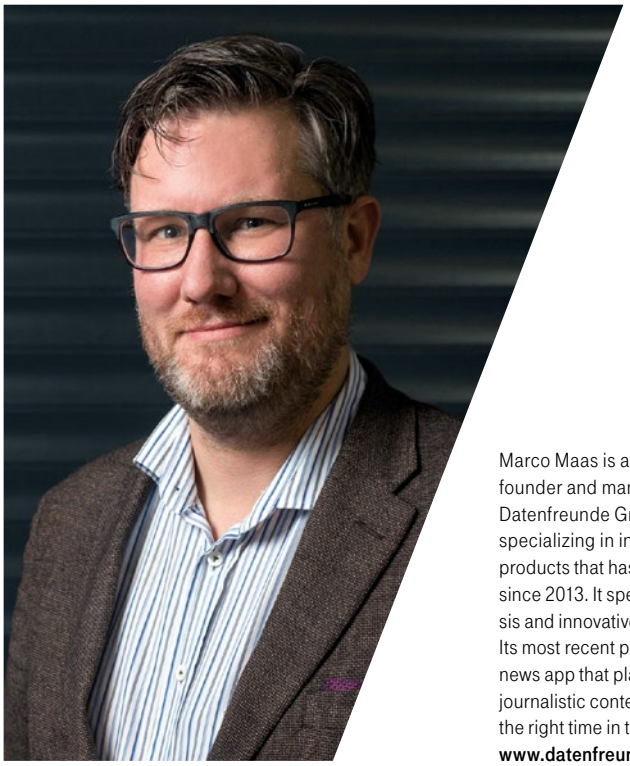
Prof. Günther Schuh holds the Chair of Production Systematics and Production Management and is Managing Director of the WZL Laboratory for Machine Tools and Production Engineering at RWTH Aachen University. He founded StreetScooter GmbH, which produces e-transporters, and e.GO Mobile AG for e-cars, where he is still CEO today.
www.wzl.rwth-aachen.de, www.e-go-mobile.com

What does real time mean in production control?
To begin with, we are talking about relative real time. And that depends on the specific requirement. There is an absolute real time created by the transmission rate of signal processing and that is in the millisecond range; however, in production control we often do not need that at all. For production planning – which is weekly – a delay of 10 to 30 minutes is typically very tolerable. For assembly control, however, it cannot exceed 30 to 120 seconds in order to be able to respond to disruptions accordingly. With cutting machine tools, on the other hand, it would be good even to find out within milliseconds when machining goes off the rails.

What requirements need to be met for this?
To start, I need to learn to see digitally. This pertains to the individual sensor: Does it work optically, with a camera using image recognition or even using radar? Then I need to learn how to interpret what is seen. This means how can I combine the sensor data into a single image, a digital shadow of production? To do this, I need to establish networks at various levels. In each machine tool, I combine sensors and actuators; at the shop floor level, I merge the various systems. How quickly I do that is also a question of telecommunication. 5G will certainly make a lot of it go faster.

What are the advantages of real time production control?
In the broadest terms, it helps to avoid the two greatest obstacles to productivity: waiting and searching. It allows me to have complete control of the control loops of my production. I know the statuses of my machines and I know where the parts are located. This affects the entire supply chain. It also increases my OEE, or overall equipment effectiveness. Now I only have productive employees and no longer need troubleshooters. When there is no more searching and waiting, the work atmosphere also improves. The employees are no longer interrupted and can work at their own individual pace. Real time capability is also important for making predictions. When I compare the current situation with past experience, I can recognize patterns and make predictions. In this way, you could predict up to 90 percent of all disruptions.

Is this kind of production control a topic for the future?
If you want to know whether this kind of control is widespread: no. If you are asking whether it can be now: absolutely! All of the necessary technology is there, it just has to be utilized. In our production of the e.GO, we have real time control with a delay of one to two minutes. And this despite the fact that a lot of assembly is done by hand. There is absolutely no reason not to immediately introduce this kind of control into every type of production. The excuse of waiting for 5G, i.e., that it does not work with 4G, is unacceptable to me.



Marco Maas is a data journalist and founder and managing director of Datenfreunde GmbH, a company specializing in innovative audio products that has been on the market since 2013. It specializes in audio analysis and innovative news distribution. Its most recent project is "xMinutes," a news app that plays out context-based journalistic content – the right news at the right time in the right place.
www.datenfreunde.de

"Two time-based drivers"

Real time in the media: Data journalist and entrepreneur Marco Maas knows the possibilities emerging for journalism and advertising.

What are the possibilities for real time applications in media?
There is a principle that states, "The better I know the customer, the better I can appeal to them." For this purpose, information from online targeting needs to be combined with information from smart home applications, motion detectors, etc.

What technology is needed for this and to what extent are such applications being used today?
From the examples mentioned, we are only one small step away. It would be necessary to consolidate collected data at various points. The technical possibilities are there and only need to be employed in the field, such as AI applications that automatically organize unstructured information and draw conclusions from it. A lot is already being done today, such as with VHF radio, where you can do things like run ads other than for grilled sausages in distribution areas that are expecting rain.

What does this mean for journalism and what does it mean for advertising?
News can be made compatible with the user and their specific usage situation. I not only know where and how the user lives, but also the situation they are in at the moment and what news they have already heard up until now. This way, regional or even local news is becoming increasingly important. Also, I can run news on a topic in which the user is already interested, or on traffic and weather before they leave the house. A similarly pointed mode of address is possible in advertising. Someone listening incessantly to Rammstein on Spotify naturally will not get an ad for a Rihanna concert. Everyone gets ads for laundry detergent, but perhaps with an emphasis on being dermatologically tested for families with a newborn child. For an even more pointed targeting, someone with a beard would then get ads for beard care products right as he is standing in front of the mirror at 7:00 a.m.

What does real time mean here and what are the response times?
In these configurations, there are basically two time-based drivers: The first is the pace set by events. Naturally, everyone wants to get breaking news that, say, Angela Merkel has resigned as chair of the CDU, as soon as possible. Local and regional topics are less time-critical. If I learn about events in my neighborhood one or two hours after the fact, that is usually perfectly fine. Exceptions might be traffic reports that I would want to see before I leave the house. The other time-based driver is created by how the user's day goes. If I want to reach minimal time windows, whether with news or ads, I need to be quick. Ideally, I know before the user that they will be interested in a certain topic in five minutes. That would then be a negative response time.

Photo: Bosch, Nico Piatkowski

LOGISTICS

“Perpetuate the flow of material”

Real time in logistics: The benefits that can be gained, especially in warehouse logistics, as explained by Dr. Nico Piatkowski, chair of artificial intelligence at TU Dortmund University.

At the TU Dortmund, Dr. Nico Piatkowski researches machine learning (ML) for resource-constrained systems. According to Piatkowski, AI and ML should not imitate people as such, but their consumption of resources. While Google’s AlphaGo did indeed defeat an international champion of the board game Go, it used 50 times the energy of its “colleagues,” who as generalists can solve much more than just one computer task.
www.tu-dortmund.de

What does real time mean in logistics?
Real time is basically a relative term. Just technically speaking, real time means the time a system needs to initiate an action within a specified period. This is in the nanosecond range. In an application, of course, I always need to consider the context and the specific requirements. A driverless delivery vehicle or a drone probably needs to make 60 to 100 decisions per second about parameters such as direction and speed – and each of these decisions needs to be correct. When I think globally, such as a container ship sailing around the world, I have days to optimize the routes used to forward goods once they arrive at the port of destination.

You specifically addressed warehouse logistics in a pilot project with one client. What is that about?
We optimized a warehouse together with an internationally renowned logistics service provider. The task was to perpetuate the flow of material from a high bay warehouse to workers at the packaging stations in order to keep the people there continuously busy. The critical thing was to come to a dynamic arrangement in the warehouse. Dynamic means, in the basic requirement for a large grocery warehouse, for example, something like placing seasonal goods – say, chocolate Easter bunnies, to be topical – in the locations in the warehouse that can be reached the fastest. But there are also much more subtle correlations that a person may not even realize. This is why we also installed a system that learns in real time and which can favorably position goods that are in high demand at certain times in variable cycles proactively.

How are learning warehouses controlled?
It could have been done with a central computer. Aside from an energy requirement that is high but, in the long run, not more economical, such a system also has the disadvantage of not being scalable with the size of the warehouse – for warehouses of different sizes, you would need central computers of different sizes. This is why we spread the problem out and gave each storage unit a minicomputer. These minicomputers do not need to compute any comprehensive, overall solution, only a partial solution. They then communicate it only to the storage units in their immediate vicinity. If I then have adequate natural light in the warehouse and can give the computers a solar solution, I end up with an autonomous self-learning system.

Are there special infrastructural requirements for this, such as 5G?
5G is more important for use on the road if I want to control driverless vehicles or delivery drones. Although I also need to integrate intelligence into the respective cars in that case. On one side, even a high-performance computer routing an entire fleet of drones reaches its limit at some point. And on the other side, a 5G connection can naturally be interrupted at times. This is why I need at least enough intelligence in a delivery drone that it can safely exit traffic if there is a problem.