The Promise of Edge Computing

Weisong Shi, Wayne State University
Schahram Dustdar, TU Wien

The success of the Internet of Things and rich cloud services have helped create the need for edge computing, in which data processing occurs in part at the network edge, rather than completely in the cloud. Edge computing could address concerns such as latency, mobile devices' limited battery life, bandwidth costs, security, and privacy.

The term Internet of Things (IoT) was first introduced to the technology community in 1999 in reference to automated supply-chain management.[1] The concept of enabling a computer to sense information without human intervention was then applied to other fields such as healthcare, home technology, environmental engineering, and transportation.

With IoT implementation now becoming more widespread, we're entering the post-cloud era, in which devices will generate a lot of data at the end of the network and many applications will be deployed at the edge to process the information. Cisco Systems predicts that an estimated 50 billion devices will connect to the Internet by 2020.[2] Some of the applications they run might require very short response times, some might involve private data, and some might produce huge quantities of data. Cloud computing can't support these IoT applications. *Edge computing*, on the other hand, can do so and will promote many new IoT applications.

WHY DO WE NEED EDGE COMPUTING?

Edge computing will become important for several reasons.

Cloud services

Moving all computing tasks to the cloud has been an efficient way to process data because there's more computing power in the cloud than in the devices at the network edge. However, although data-processing speeds have risen rapidly, the bandwidth of the networks that carry data to and from the cloud hasn't increased appreciably. Thus, with edge devices generating more data, the network is becoming cloud computing's bottleneck.

As an example, cameras in an autonomous vehicle capture a huge amount of video data, which the system must process in real time to yield good driving decisions. If the vehicle must send the data to the cloud for processing, the response time would be too long. And a large number

0018-9162/16/$33.00 © 2016 IEEE

of autonomous vehicles in one area would further strain network bandwidth and reliability.

Processing data at the network edge would yield shorter response times, more efficient processing, and less pressure on the network. Recent work on micro-datacenters (mDCs; small, modular datacenters designed to optimize networked devices' performance via the cloud)[3] and cloudlets (small cloud datacenters at the Internet's edge designed to work with mobile devices)[4] has studied this.

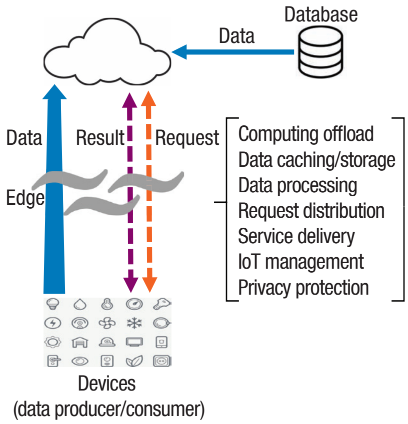

Figure 1. In edge computing, the cloud collects data from existing databases, as has been done traditionally, and also from end devices such as sensors and mobile phones. The devices act as both data consumers and data producers. Thus, requests between end devices and the cloud are bidirectional, instead of just from end devices to the cloud as in the past. Nodes at the network edge perform many computing tasks—including data processing, caching, device management, and privacy protection—to reduce traffic from devices to the cloud. IoT: Internet of Things.

The IoT

Billions of electrical devices—as well as other devices such as air-quality sensors, LED bars, and streetlights—will become part of the IoT and will produce, as well as consume, data. Conventional cloud computing won't be efficient enough to handle the sheer volume of data they'll generate. For example, the Cisco Global Cloud Index estimates that by 2019, people, machines, and things will produce 500 zettabytes (a zettabyte is 10^{21} bytes) of data but that global datacenter IP network traffic will reach only 10.4 zettabytes.[5]

Typically, data producers generate raw information and transfer it to the cloud, and data consumers send requests for information to the cloud. However, this structure won't work with the IoT because of the large data volumes involved.

Using cloud computing with the IoT will also raise concerns about the privacy of transferred data. In addition, most of the IoT's end nodes are power constrained. Offloading some computing tasks to the network edge could be more energy efficient. The OpenFog Consortium's proposed fog computing paradigm[6]—an infrastructure in which some application services are handled by devices at the network edge and some by a cloud-based datacenter—would enable this.

From data consumer to data producer

In cloud computing, devices such as mobile phones at the network edge traditionally only consume data, such as by enabling a user to watch a video. Now, though, users are also producing data with their mobile devices, such as by uploading posts and photos to social-networking sites.

This change requires more functionality at the network edge. For example, many people share images or videos via a cloud-based social-networking service such as Facebook, Twitter, or Instagram. However, uploading a large image or video clip could require a lot of bandwidth. In this case, a device at the network edge could reduce the video clip's resolution, and thus its size, at the edge before uploading it to the cloud. In another example, properly equipped edge devices could better protect privacy by processing sensitive data collected by wearable personal health devices, rather than uploading the data over a vulnerable network to the cloud.

WHAT IS EDGE COMPUTING?

Edge computing, shown in Figure 1, refers to the enabling technologies that allow computation to be performed at the network edge so that computing happens near data sources. It works on both downstream data on behalf of cloud services and upstream data on behalf of IoT services. An edge device is any computing or networking resource residing between data sources and cloud-based datacenters. For example, an edge device could be a smartphone sitting between body sensors and the cloud, or an mDC or cloudlet between a mobile device and the cloud.

In edge computing, the end device not only consumes data but also produces data. And at the network edge, devices not only request services and information from the cloud but also handle computing tasks—including processing, storage, caching, and load balancing—on data sent to and from the cloud. The edge must be designed well enough to handle such tasks efficiently, reliably, securely, and with privacy in mind. It thus must support requirements such as differentiation, extensibility, isolation, and reliability.

Edge computing could yield many benefits. For example, researchers have shown that using cloudlets to offload computing tasks for wearable cognitive-assistance systems improves response times by between 80 and 200 ms[4] and reduces energy consumption by 30 to 40 percent. CloneCloud technology reduces response times and power usage by 95 percent for tested applications, in part via edge computing.[7]

Edge-computing practitioners should be aware of several issues, including system reliability. Also, energy-constrained edge devices could fail because of short battery life or inadequate wireless communications. In these cases, developers should still enable the system to provide its basic functions.

Security and privacy are additional concerns. On one hand, edge computing better protects data because processing occurs closer to the source than in cloud computing. However, supporting security and privacy is more difficult in edge computing due to network topology, the many inexpensive personal mobile devices in the system, and sensor unreliability.

IMPLEMENTATION EXAMPLES

Two examples demonstrate practical edge-computing implementations.

Online shopping

Online shopping could benefit from edge computing. For example, a customer might make frequent shopping-cart changes. Traditionally, these changes occur in the cloud, which then updates the shopping-cart view on the user's device via the network. This process might take a long time depending on network speed and server loads. Delays might be even longer for mobile devices because of wireless networks' relatively low bandwidth.

Cloud computing itself could also cause latency, which diminishes the user experience. Avoiding this is important because shopping with mobile devices is becoming increasingly popular. Caching shopping-cart data at the edge and offloading shopping-cart updates from cloud servers to edge nodes dramatically reduces latency. Numerous research projects have addressed how this type of offloading enhances performance and reduces energy consumption in a mobile-cloud environment.[7–10]

Data at the edge node could subsequently be synchronized with the cloud in the background.

Finding a missing child

Edge computing could help officials find missing children. Today, cameras deployed in public areas in cities—as well as cameras in some vehicles—could capture a missing child's image. However, this frequently isn't leveraged because the camera data usually isn't uploaded to the cloud due to privacy concerns or the cost of transferring the information. Even if the images are in the cloud, accessing and searching such a huge quantity of data could take a long time, which isn't acceptable when looking for a missing child.

With edge computing, the data could be pushed to the many edge devices in a target area. They could search the data they receive and report the findings to the cloud, yielding the results much faster than using only cloud computing.

OPEN ISSUES

Edge computing needs killer apps to reach its potential. Several application domains are worth exploring such as disaster response and management, body cameras for police officers, smart vehicles, and connected health systems.

Programmability

In edge computing, as several research projects have demonstrated, programmers must partition the functions of their applications between the edge and the cloud.[7–9] Most early efforts in this area were done manually and carefully tuned, which isn't scalable or extensible. Thus, easy-to-use programming frameworks and tools are required.

Naming

There are many edge devices and applications. Currently, however, there's no efficient, standardized naming system for edge-computing devices and applications so that they can be easily found. Thus, edge practitioners usually must learn many communication and network protocols to communicate with their systems' heterogeneous elements. Edge computing needs a naming scheme that will handle device mobility, highly dynamic network topology, and privacy and security protection, while also enabling scalability.

Privacy and security

The network edge presents security and privacy challenges. For example, a hacker could learn a lot by reading data going to and from a smart-home system. By capturing electricity or water usage, the hacker could easily determine if the house is probably vacant and thus vulnerable to burglary. A lack of efficient tools is one of the challenges to protecting data security and privacy at the network edge.

Edge computing could scale from a single person to a smart home to even an entire city. Given that a city with 1 million people will produce an estimated 180 petabytes of data per day by 2019,[5] the benefits could be enormous.

However, to realize this vision, the systems, network, and application communities must work together, joined by the many groups that could benefit from the technology such as those in environmental and public health, law enforcement, fire protection, and utility services.

In the past few years, this process has begun. For example, proponents formed the OpenFog Consortium (www.openfogconsortium.org) in November 2015 to promote an ecosystem to accelerate the adoption of open fog computing by bringing together companies, universities, and individual researchers. In addition, new conferences are planned, such as the IEEE/ACM Symposium on Edge Computing (SEC), to be held in October 2016; and the Mobile Edge Computing Congress (MEC), to be held in September 2016. If this trend continues, edge computing will be on its way to fulfilling its promise.

REFERENCES

1. K. Ashton, "That Internet of Things Thing," *RFID J.*, vol. 22, no. 7, 2009, pp. 97–114.
2. *The Internet of Things: How the Next Evolution of the Internet Is Changing Everything*, white paper, Cisco Systems, Apr. 2011.
3. A. Greenberg et al., "The Cost of a Cloud: Research Problems in Data Center Networks," *ACM Computer Communication Rev.*, vol. 39, no. 1, 2008, pp. 68–73.
4. M. Satyanarayanan et al, "The Case for VM-Based Cloudlets in Mobile Computing," *IEEE Pervasive Computing*, vol. 8, no. 4, 2009, pp. 14–23.
5. *Cisco Global Cloud Index: Forecast and Methodology, 2014-2019 White Paper*, white paper, Cisco Systems, 2014.
6. F. Bonomi et al., "Fog Computing and Its Role in the Internet of Things," *Proc. 1st Ed. MCC Workshop Mobile Cloud Computing* (MCC 12), 2012, pp. 13–16.
7. B.-G. Chun et al., "CloneCloud: Elastic Execution between Mobile Device and Cloud," *Proc. 6th Conf. Computer Systems* (EuroSys 11), 2011, pp. 301–314.
8. A. Rudenko et al., "Saving Portable Computer Battery Power through Remote Process Execution," *ACM SIGMOBILE Mobile Computing and Communications Rev.*, vol. 2, no. 1, 1998, pp. 19–26.
9. G. C. Hunt and M.L. Scott, "The Coign Automatic Distributed Partitioning System," *Proc. 3rd Symp. Operating Systems Design and Implementation*, 1999, pp. 187–200.
10. S. Kosta et al., "ThinkAir: Dynamic Resource Allocation and Parallel Execution in the Cloud for Mobile Code Offloading," *Proc. 31st IEEE Int'l Conf. Computer Comm.* (Infocom 12), 2012, pp. 945–953.

WEISONG SHI is a professor of computer science at Wayne State University, where he heads the Data Science Initiative. Contact him at weisong@wayne.edu.

SCHAHRAM DUSTDAR is a professor of computer science at TU Wien (Vienna Technical University), where he heads the Distributed Systems Group. Contact him at dustdar@dsg.tuwien.ac.at.