July 20, 2018

Distributed Placement of Machine-Learning Computing in an Edge Network

Karthik Ravi Shankar

Gurunathan Nagarajan

Distributed Placement of Machine-Learning Computing in an Edge Network

Abstract:

Smart devices continue to proliferate as the Internet-of-Things expands. Collectively, Internet-of-Things devices generate massive amounts of data for processing, analysis, and implementation. However, most individual smart devices lack sufficient hardware resources to process collected data in an efficient or timely manner. Thus, most devices send their data to a remote server or other cloud-based computing system for processing because of the increased computational capacities of such remote locations. Although these remote locations can process the data faster and more efficiently, the increase in the number of smart devices accessing the remote locations increases the transmission traffic, and associated bottlenecks, on networks and other data-transmission systems. Many smart devices reside on local networks that feature other, more-powerful, computing devices, such as desktops, laptops, home servers, and gaming systems. Some of these additional computing devices could be tasked with processing data and other information for Internet-of-Things devices that lack sufficient computational capacity to process the data themselves.

Keywords:

Smart device, WiFi, Bluetooth, Internet-of-Things, edge, data processing, machine learning, smart home, automation, cloud, security, latency.

Background:

As the Internet-of-Things expands, smart devices continue to proliferate. Internet-connected thermostats, appliances, vehicles, lights, cameras, robots, and machines are found in all areas of life, including home, work, business, recreation, and school. In particular, many smart devices leverage machine learning algorithms, models, and processes to sense the environments in which they are located and respond to changes in or conditions of those environments. Although these processes can be run on the smart devices themselves, many smart devices do not include sufficient resources to efficiently process the collected information and must transmit the information and data to other computing devices, which run the machine learning processes for the smart device. For example, smart devices may transmit information to the cloud or a remote server for processing. As the number of smart devices increases, the load on transmission networks also increases. Additionally, each transmission of information increases the opportunity for data corruption, interception, or loss. Also, processing the information at a remote location could implicate matters of security or privacy.

Description:

Smart devices continue to proliferate as the Internet-of-Things expands. Collectively, Internet-of-Things devices generate massive amounts of data for processing, analysis, and implementation. However, most individual smart devices lack sufficient hardware resources to process collected data in an efficient or timely manner. Thus, most devices send their data to a remote server or other cloud-based computing system for processing because of the increased computational capacities of such remote locations. Although these remote locations can process the data faster and more efficiently, the increase in the number of smart devices accessing the remote locations increases the transmission traffic, and associated bottlenecks, on networks and other data-transmission systems. Many smart devices reside on local networks that feature other, more-powerful, computing devices, such as desktops, laptops, home servers, and gaming systems. Some of these additional computing devices could be tasked with processing data and other information for Internet-of-Things devices that lack sufficient computational capacity to process the data themselves.

Smart computing devices include things like thermostats, home assistants, ovens, refrigerators, cameras, lights, switches, security systems, monitors, robots, appliances, mobile phones, tablets, or home theater components. A smart device can be directly connected to the Internet or indirectly connected to the Internet via a communication hub. The smart devices can be accessed and manipulated using a web browser, an application on a mobile device, a local remote, or programing system (*e.g.*, voice commands, home automation systems). The smart devices can also connect and interact with one another. For example, a home theater system that is turning on could send a signal that instructs the lights in the room to dim after a certain amount of time or a security camera could recognize an authorized individual and automatically open a door lock.

For example, Figure 1 shows smart devices in a kitchen, such as a thermostat, home assistant, slow-cooker appliance, oven, microwave, and refrigerator. Although some devices perform computation and data analysis themselves, most of these smart devices connect to a remote server or other cloud-based computing system to process the data collected by the smart device in order to carry out its respective smart functions, as illustrated by the cloud connections of each device. These smart devices may connect to the Internet individually via any number of wireless or wired networks. The smart devices may also connect to one another over Bluetooth, Z-wave, Zigbee, or other local, wired, mesh, or wireless networks.

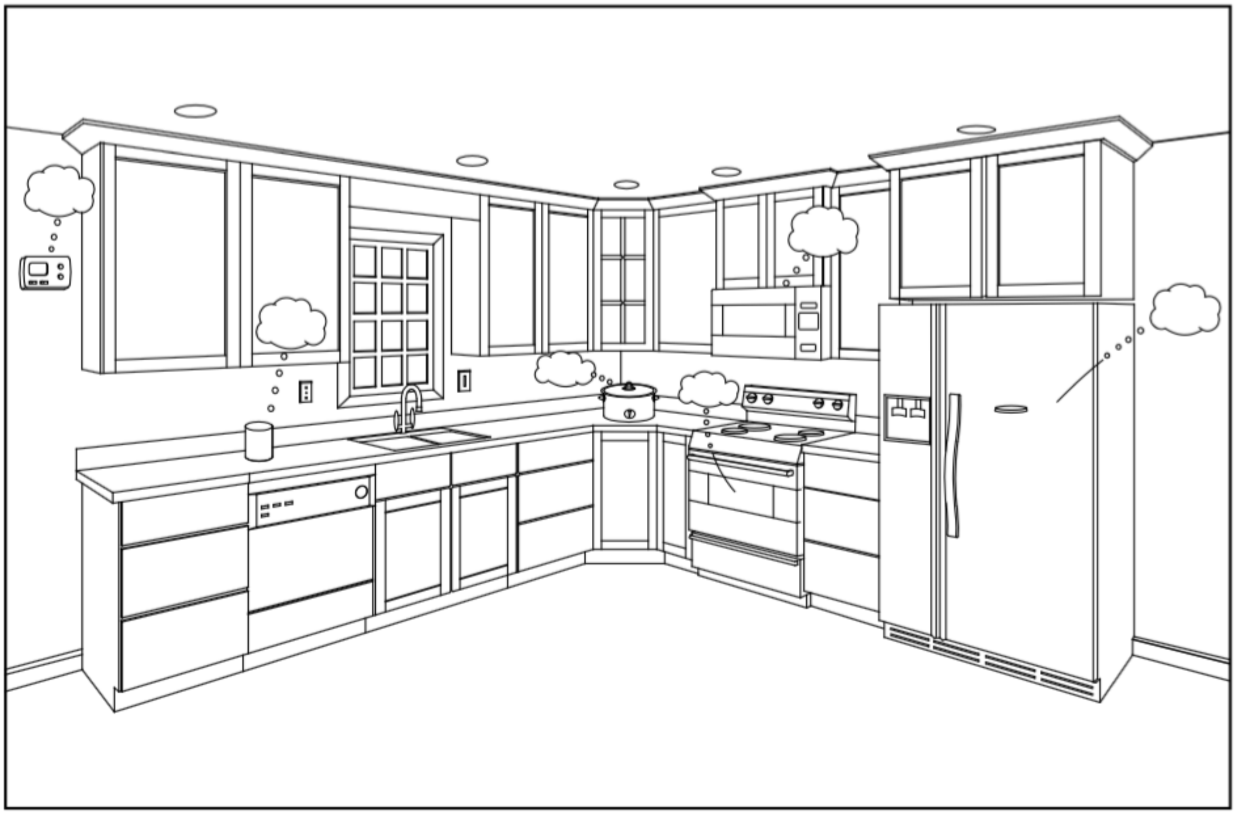

Figure 1

Each of the smart devices shown in Figure 1 may be able to perform certain operations without the need of significant computation or data analysis. For example, the smart thermostat could execute a scheduled temperature profile or pattern without need of connecting to the internet or processing significant data. However, the thermostat could also be equipped to monitor the temperature in the kitchen throughout the day and proactively manage the temperature. For example, the smart thermostat could correlate operation of the oven with an increase in temperature in the room and may interpret this correlation as baking. The thermostat could either respond to the oven being turned on or could anticipate the oven being turned on at a particular time based on a previous historical pattern. Such monitoring, connecting with other devices, and inference making or prediction often requires more computational ability than the thermostat alone can handle. The edge system described herein can distribute the computational load to other local devices instead of transmitting the information over the internet to a remote server or other cloud-based computing system.

Figure 2 illustrates several different computing devices that generally have more computational capacity than many Internet-of-Things devices. Figure 2 includes a desktop computer, a tablet computer, a smartphone, a laptop computer, a home server, a gaming console system, a television, and a positionable tabletop computer. Any number of other hubs, computing systems, and electronic devices common to homes and businesses could also have sufficient capacity to conduct the computational analysis of data collected by less-capable smart devices.

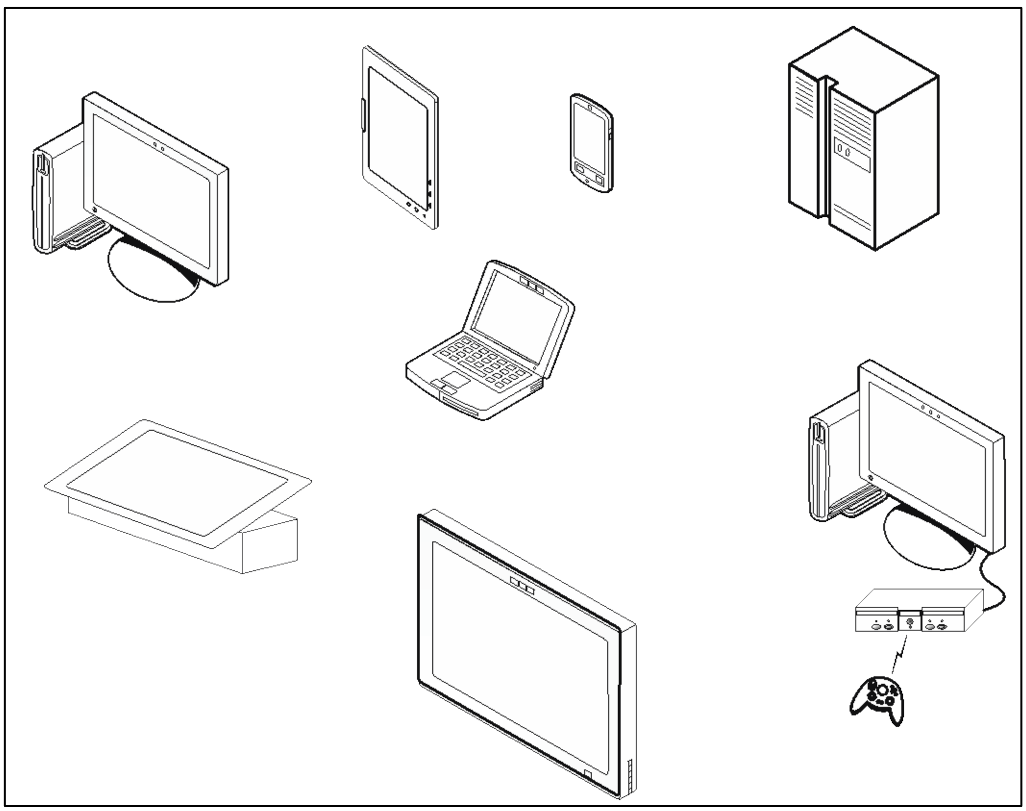

Figure 2

The local edge data-processing system consists of three main parts. First, a cross platform application programming interface that is compatible with common operating systems on the more-capable computing devices, such as those of Figure 2, and capable of communicating with the local smart devices. Second, the system executes a task planner. The task planner collects information from the more-capable computing devices regarding their individual computational capabilities, current or scheduled computational loads, and available computing resources as well as estimated costs of computation. The task planner could account for various computational costs in determining where, when, and how to distribute the computational processing operations. Third, the system distributes the computation processes securely, in separate containers, to the various more-capable computing devices. In this manner, different machine-learning operations of various smart devices can be kept separate, which can increase security, decrease noise and possible cross-contamination, and increase efficiency. Optionally, the system can include a dedicated edge-computing computational block that can manage or multiplex multiple workloads among the various devices. In concert with the task planner, this edge-computing computation block could assist with determining whether or not some processing should bypass the edge network and be transmitted for processing at a remote server or other cloud-based computing system.

Figure 3 shows an edge network in operation. The smart devices located in the kitchen of Figure 1 and the more-capable computing devices of Figure 2 reside in the house. All of these devices are connected to one another by wired or wireless network connections.

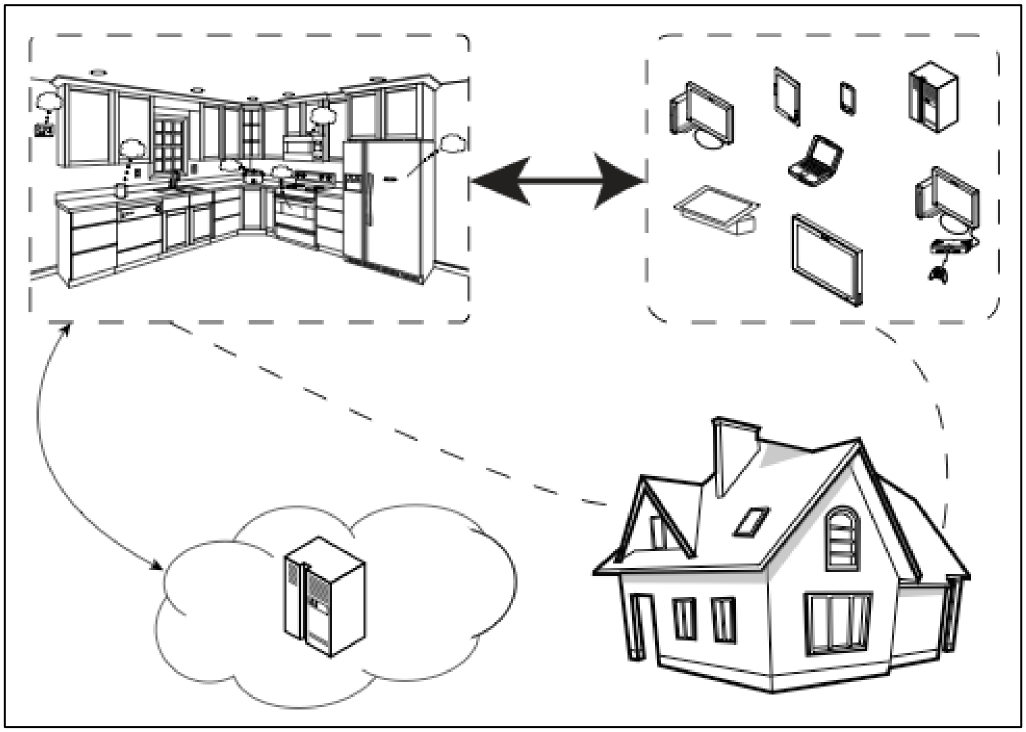

Figure 3

The edge network system described here analyzes the various computational requirements or needs of each of the smart devices. When the computation needs exceed the capacity of the smart device, the system can transfer that computational load to any one of the more-capable computing devices, such as a desktop computer. The system accounts for the various loads currently on the more-capable devices and balances that against the constraints on the smart device (*e.g.*, balancing the need of the smart device to receive its operational instructions derived from the information collected against the capabilities of the more-capable computing device). If the system determines the need of the smart device outweighs the capability of the edge network system to process the data, the system may transmit the data to a remote server or other cloud-based computing system. However, as illustrated by the size of the double-ended arrows, the majority of the data processing occurs locally.

Local data processing in an edge-network system has several distinct advantages. Local data processing decreases the transmission load on internet data transmission channels. As the number of smart devices increase, the traffic on data transmission channels also increases, which can result in slow downs, bottlenecks, and even data corruption. Some may even have to purchase faster internet connections to be able to have the data from smart devices processed quickly enough. Data transmitted over internet data channels also increases potential security risks. Retaining the data locally decreases transmission security risks. An edge network optimizes local resources and computing devices and can decrease the costs of cloud computing.

Although remote locations with increased computational capacity often can process the data faster and more efficiently, many smart devices reside on local networks that feature other, more-powerful, computing devices, such as desktops, laptops, home servers, and gaming systems. Some of these additional devices could be tasked with processing data and other information for Internet-of-Things devices that lack sufficient computational capacity to process the data themselves.