element14
AN AVNET COMPANY

Smart sensors – enabling the Intelligent Internet of Things

element14
AN AVNET COMPANY

Contents

Abstract

Engineers working on systems built using Internet of Things (IoT) technologies now have access to a huge range of sensor options. The available modalities have moved well beyond temperature and pressure to encompass sensors that can act as the eyes, ears and noses of the IoT. Thanks to the connectivity made possible by the IoT, engineers no longer need to incorporate all of these sensors into their equipment. Open IoT protocols make it possible to exchange data with sensors distributed around a network that are built into other devices as long as the systems that communicate with them have the right credentials. In turn, advanced sensor-fusion techniques provide the tools for the inputs from many different types of input to be incorporated into a single smart model.

This white paper describes the choices available to engineering teams and the things to consider when selecting and integrating sensors into their IoT-driven systems and applications.

Introduction

Sensors are the eyes and ears of the Internet of Things (IoT). With novel chemical and spectroscopic technologies, they are even becoming the noses of the IoT. Sensors already provide the real-time raw data to a growing range of applications that use it to produce insights. A major current change is the introduction of smarter sensors - devices that incorporate value-added functions to ease integration into IoT networks and provide valuable information more readily.

Key trends in sensor development include:

- a focus on low power consumption so that the devices can be used for years on a single battery charge or even harvest energy from the environment
- the ability to connect sensors using wireless networks even over distances measured in kilometres
- the inclusion of security protections to prevent unwanted reprogramming or malign use of sensor data
- the use of data-fusion techniques to provide better analyses of changes in the environment captured by multiple sensor types

This white paper analyses the choices on offer and outlines how integrators can take best advantage of the new generation of smart sensors.

Key developments in sensor modalities

There are many sensor options open to systems integrators and electronics systems designers. The possibilities multiply when the power of the IoT is taken into account. In traditional electronics design, the system's functionality needed to be self-contained. Sensors and other peripherals would be built into the core hardware, using low-level proprietary interfaces such as simple serial or analogue I/O. The only sensor information available to the system would be from these built-in devices.

IoT technology massively expands the range of sensing options for any system. With the ability to connect over a wired or wireless network, a system can, as long as it has the right credentials, obtain relevant information from any sensor on the network.

This enables novel applications across a range of situations, from smart cities through building automation to agriculture.

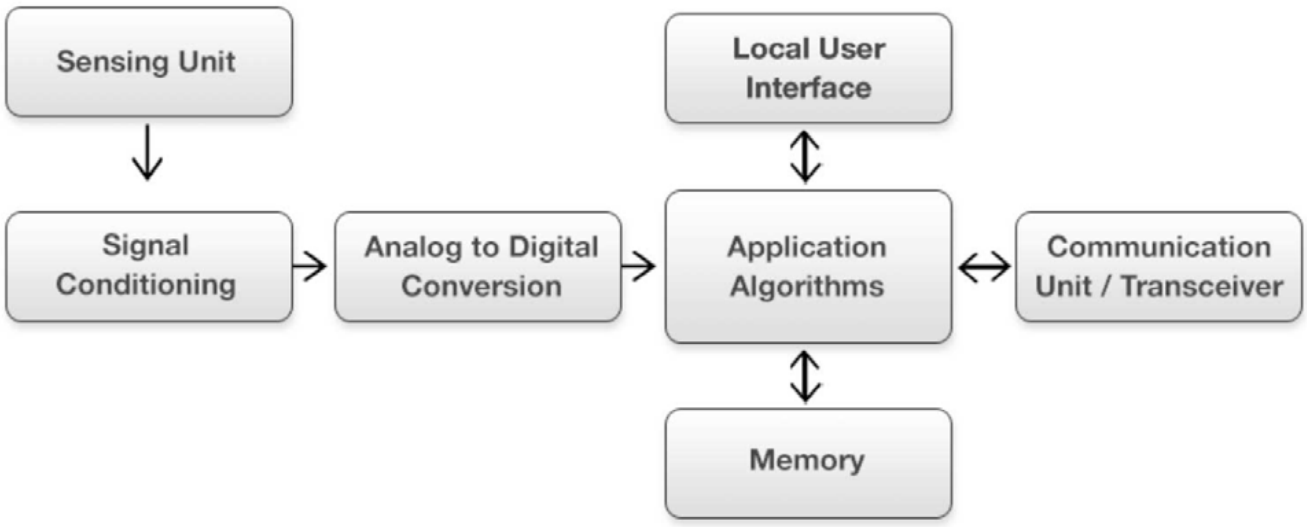

Figure 1: Smart sensor building blocks

To look at an example, in building automation as well as in smart cities, an important task is people counting. This information can inform many applications, from heating and lighting to advanced intruder-detection systems. Traditional approaches to automated lighting in offices use comparatively simple motion detectors, using passive infrared (PIR) detection or similar technologies. Though these simple systems have been in use for many years, they are prone to errors in an office environment where people may not move around enough to prevent the software from deciding that the room is unoccupied, and inconveniently turning off the light.

The new wave of people-counting systems can make use of a range of sensors to avoid the mistakes made by simpler systems. Not only can the application take advantage of different sensor modalities, it can also harness sensors in different positions around and outside a room. The system can use that information to determine how people are moving around and confirm what information other sensors are reporting. An individual sensor need not provide information only to this one application. As well as controlling lighting, data from cameras and other sensors can support heating, air-quality and security applications.

Among the first electronic sensors to be made available were temperature-measurement devices as well as pressure sensors developed for automotive engine control and simpler forms of motion and proximity detector employed in industrial control. The automotive sector has been responsible for reducing the cost of many different types of sensors. One example is the ultrasonic sensor, which was developed initially as a parking assistance technology. Sensors mounted in front and rear fenders or bumpers emit sound waves well outside the human range of hearing, using the reflection time to measure the distance between the sensor and the reflecting object. This kind of sensor is now being used to support another application of automotive parking, detecting when bays in a managed car parking facility are free.

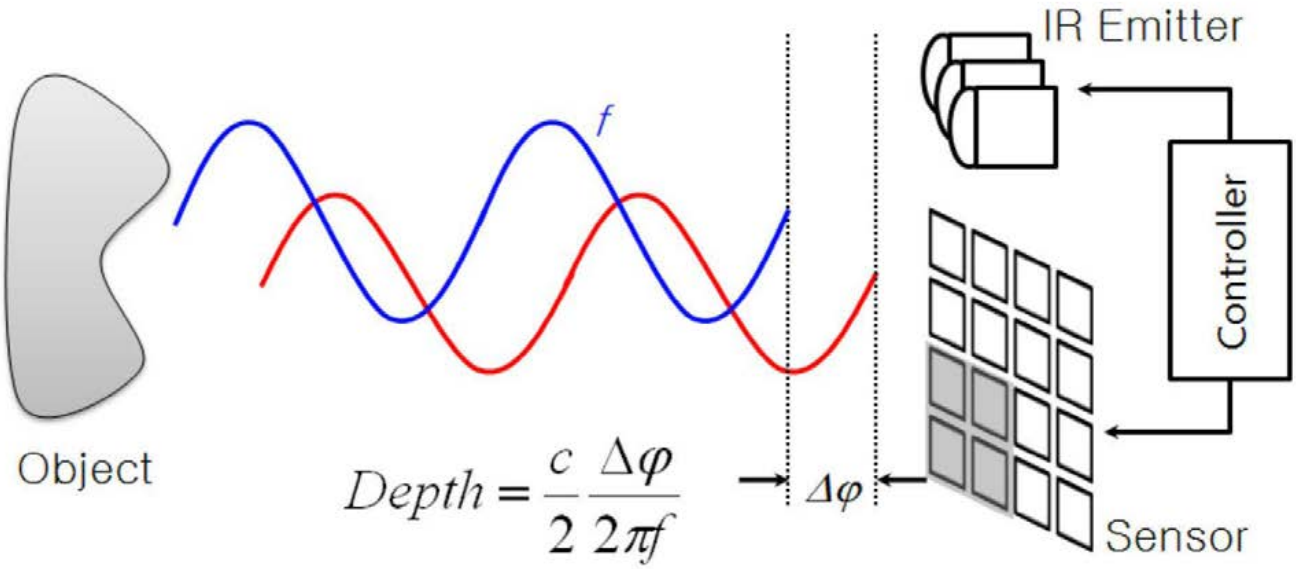

Figure 2: The principle of ToF depth camera [37,71,67] : The phase delay between emitted and reflected IR signals are measured to calculate the distance from each sensor pixel to target objects.

Some of the most dramatic advances in technology have been in the field of image and 3D sensing, some of it driven by high demand from the consumer and automotive sectors. Visible-light camera modules are among the most differentiated options, with a wide variety of resolutions and lens types available from multiple vendors. Many are provided in the form of imaging modules, some designed for direct connection to a System on a Chip (SoC) though digital serial interfaces. Others are smart modules that have built-in processing modules to support security and building-management applications.

2D sensors provide valuable information to autonomous systems such as robots, not just when they are embedded into the machine itself but also when positioned in the surrounding environment. By combining imagery from external sources, the robot can gain a far better understanding of obstacles and objects it cannot see directly and plan its movements accordingly. Sensing technologies such as radar, LiDAR and time-of-flight imaging provide even more information by giving systems the ability to perceive objects in 3D space instead of having to project 2D images into a virtual 3D world model.

Radar and LiDAR operate using the same core concept. Like ultrasonic sensors, they use reflections from objects to determine where they are and how far away, albeit with much greater resolution. The key difference between the two forms is that they use different parts of the electromagnetic spectrum. Radar generally uses wavelengths in the millimetre range and LiDAR operates in the infrared region, which provides higher spatial accuracy.

Another route to depth mapping is a time-of-flight (ToF) sensor. Similar to LiDAR, it uses reflections of light generated by infrared lasers to determine how far away objects are from the point of origin. The chief advantage of the ToF technology is that the sensing can be performed by a conventional RGB camera sensor, which can help reduce cost in less mission-critical applications such as people counting in a room or space.

A key advantage of the ToF camera in building management is that it can support accurate people-counting analysis in a privacy-aware fashion. The higher accuracy and more specialised sensors used for LiDAR may be preferred for robotic control.

Though imaging sensors are important, IoT applications call for other modalities. Acoustic sensors can be used in concert with imaging technologies to augment a system's ability to perform security monitoring. The MEMS microphone of an acoustic sensor can pick up the characteristic noise of a window breaking. Systems that analyse frequency and transients can readily distinguish between such an event and normal background noise.

In buildings, the importance of air quality was brought into sharp focus by the Covid-19 pandemic. Even before the virus began to spread, air quality and the concentration of gases such as carbon dioxide and volatile organic compounds were becoming a concern. The availability of fresh air has implications for health and productivity. Recent research has demonstrated a link between aerosol and carbon dioxide levels such that if high levels of the gas are detected, air-conditioning should be engaged to bring in more fresh air in order to reduce the concentration of airborne virus and other particles.

A number of chemical-sensor technologies are available. Many are focused on single gases but some more sophisticated spectroscopic sensors that are emerging can detect multiple gases and chemicals. A new generation of photo acoustic sensors is also making widespread air-quality measurement easier to implement. These sensors couple an infrared emitter with an optical filter tuned for the target gas. In high concentrations, absorption by the gas molecules triggers a change in pressure that is detected by the acoustic element.

These sensors join many others in enabling a growing range of sophisticated applications, many of which will combine data inputs from several sources.

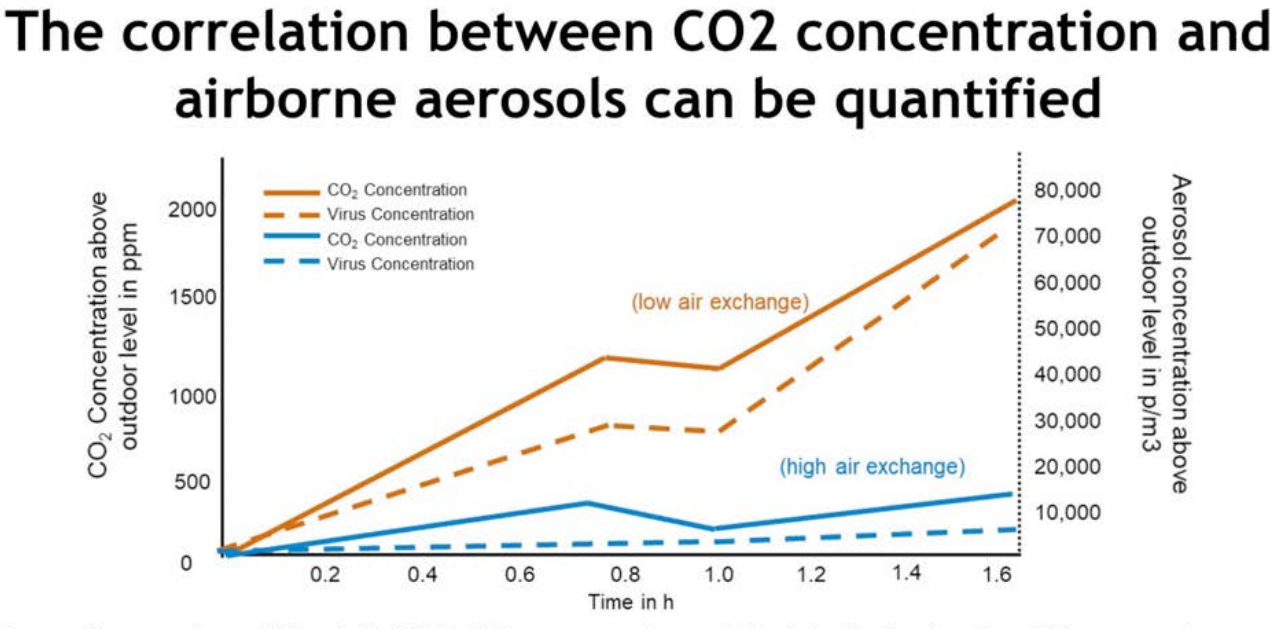

Figure 3: Risk assessment of aerosols loaded with virus based on CO2 concentration
Source: Hartmann, A., and Kriegel,M . (Berlin Institute of Technology, 2020)

Read more on: Integration of 2D and 3D Image Sensing in IoT

Sensor connectivity options

Thanks to the IoT, the sensing range of an individual system can extend way beyond any directly connected hardware. Network connectivity provides the ability to obtain data from any sensor to which it has legitimate access. In doing so, the system can combine the data streams from the many sensor inputs to build a far more reliable picture of its surrounding environment than simply relying on its own direct inputs. How the devices are connected to the network therefore becomes an important factor in the design process.

Though the IoT is designed to let devices communicate using a common set of protocols, principally based on the TCP/IP stack, there are many differences in how they access the network, which in turn affects cost, performance, energy efficiency and reliability. The fastest network connections are often through wired connections. Ethernet is today the most widely used wired-network standard and can support datarates well into the gigabit per second range. This supports the delivery of video information at high frame rates and resolution. A further advantage of Ethernet is that it can deliver power to devices as well as data connectivity. For this reason, Power over Ethernet (PoE) is commonly used to support security-camera installations. Thanks to IoT technologies, these same cameras can feed information to building-management systems over the same Ethernet infrastructure.

Despite the advantages of Ethernet, there will be many situations where it is impractical to lay cables to all the devices that need to communicate with the core network. This is where wireless connectivity is critical. Integrators are faced with a range of options for wireless networks, ranging from high-bandwidth protocols operating at 5GHz or above, such as WiFi, down to sub-gigahertz protocols that include Z-Wave and 802.15.4-based protocols such as Wireless HART. Each has its own advantages and disadvantages.Typically, the lower-frequency protocols support low-energy, low-datarate connectivity. High-bandwidth demands will be handled at 5GHz and above..

The first consideration is one of range. Conceived as a body-area network, Bluetooth was designed for comparatively short distances though it can easily support an operational range of 100m, making it suitable for indoor networking in homes, offices, shops and warehouses. Like many IoT wireless protocols that are aimed at similar applications, Bluetooth uses the 2.4GHz unlicensed band.
To extend its potential range, more recent versions of Bluetooth have embraced the concept of mesh networking, a scheme that was implemented in a number of other IoT wireless network protocols such as Thread, Wirepas, Zigbee and Z-Wave.

Traditionally, whether wired or wireless, networks mostly employed a star topology. With this approach, all messages are relayed through a central hub or gateway. The hub is responsible for passing on each message to the actual destination for the message, which may be another device on the local star, or another star managed by a different hub elsewhere on the core network.

Mesh networking makes it possible for any device to pass on messages that it detects to any other that is in range. There is no need to communicate directly with a hub. This has an important benefit when it comes to determining the maximum effective range of a network. As long as another compatible device is in range, a node can communicate with the rest of the network even if it is out of range of the nearest hub.

A further benefit of mesh networking is easier installation. This is because the protocols need to contain some inherent level of self-management, to avoid packets being lost or routed in infinite loops with no prospect of reaching their intended destination. By contrast, star networks often need some level of management to ensure that routing tables are set up correctly and to denote one device as a hub that can take responsibility for the devices it interfaces with.

A mesh network needs to be fluid and adaptive. In most mesh protocols, an activated device broadcasts a message to alert all the nodes in range. When each node responds, devices that hear the message add the address to their own database of neighbours. Regular messages update this information - if a device drops off the network, the other devices do not try to reach it, ensuring holes do not develop in the network unnecessarily. Using the self-built tables, each message is passed along by intermediate nodes until it reaches its intended destination.

An important consideration is the activity of different nodes and tuning the radio-access mechanism for them. As packet transmission and active listening consume energy, an important trade-off in network design is the activity level of each device and the amount of data it is expected to send and receive. A potential issue for battery-powered nodes is that if they are called upon to relay messages that are not meant for them, they will use more energy than expected.

Mesh networks tend to be favoured in situations where data-transfer rates are comparatively low and where multicast or broadcast messages are commonplace. For example, in lighting control, it is useful to be able to control multiple devices with one message. The rate of updates is also comparatively low. A device used for security monitoring, however, may be expected to provide frequent updates if it detects an exceptional condition. Unless range is an issue, this will favour deployment using a star-based network topology and protocol. Over long distances of several hundred metres, low-power wide-area networks (LPWANs) are favoured. Wireless networks such as LoRaWAN and Sigfox make it possible to support devices that may be several or even tens of kilometres away from their nearest hub. Typically, these networks will handle relatively low data transfer rates. LoRaWAN, for example, supports transmission at up to 50kb/s. Sigfox operates at hundreds of bits per second and is intended for devices that need to be able to send short messages relatively infrequently. A further consideration is that Sigfox is unidirectional, enabling devices send messages but not to receive them. A backup network may be needed to support functions such as software updates, though it may be feasible to use a short-distance protocol such as Bluetooth to allow service engineers to visit the location of each sensor periodically instead of having to implement a second long-distance network interface.

LoRaWAN supports bidirectional communication and allows users to operate their own gateways or call upon the services of public operators. The arrival of 5G New Radio (NR) provides another set of choices. Though cellular connectivity has been available for IoT-class devices for many years, 5G has specific support for the low-energy demands of these applications as well as pricing models. The technology also supports relatively new types of service such as Wirepas, a mesh-based 5G-class network intended for IoT systems that can work independently of the cellular network.

The many options for network connectivity may seem overwhelming but the availability of these choices make it possible to optimise connectivity for devices and systems no matter how far away the sensors are from the applications that consume their data.

Read more on: Intelligent Devices with Smart Sensors and Sensor Fusion

Smarter sensors

The distributed nature of sensing architectures in the era of the IoT and the use of networking to integrate devices calls for them to be smarter. In-system sensors often use analogue or digital serial interfaces to send data to a host microcontroller or microprocessor. Any pre-processing or filtering of the data is performed in the host. To interface with a wireless or wired network, an IoT device will often include its own microcontroller to manage access to the network. The availability of that processing core may provide additional capacity to handle functions for secure transfers and data pre-processing and filtering, turning an IoT-compatible sensor into a smart sensor.

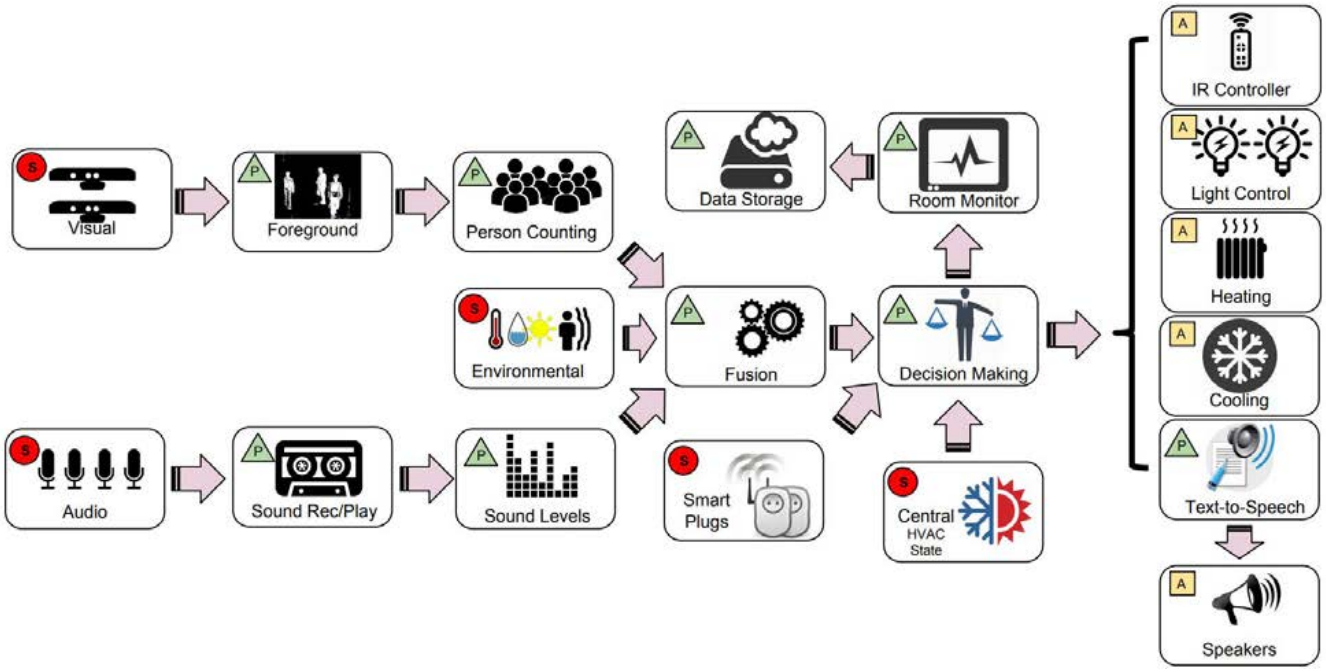

Figure 4: High-level architecture of the smart room application

Data filtering is an important feature of IoT applications, particularly in situations where available bandwidth or energy are limited. Although local processing does require some energy, using it to restrict how much data is relayed will often be a better trade-off than sending every sample over a wireless link. An additional benefit of filtering is that it reduces network load, which may be important when using a mesh network topology. For sensors connected over LPWANs, the cost of data transfer may also be an important factor in the degree of pre-processing the smart sensor needs to perform.

Filtering can take many forms. A common technique is to use thresholds to determine whether the data received shows an appreciable change. Any remote systems will maintain a cache of the last value received and assume that there is no change - only if the input goes past a limit or deviation from the last value does it send an update.

The filtering may also distinguish between changes that need action and those that are useful for updating models, but which do not call for a real-time update. This can be achieved using another set of thresholds or a local data model that determines whether the input is moving out of range. Updates that do not need to be relayed in real-time can be held temporarily in a buffer and then sent with later measurements in a single data packet.

The use of compression technologies such as linear predictive codes can further optimise the use of network bandwidth. Often the changes that need to be sent are numerically close to each other, a property that linear codes can use to reduce the number of bits needed for each sample

The situation is more complex for 2D and 3D sensors, not least because the amount of data they process is much more extensive than with 1D measurements such as temperature or pressure. Cameras developed for security, for example, may incorporate basic AI models or algorithms that examine each successive frame for changes. Small changes may be ignored. Larger changes that indicate, for example, the arrival of a person or vehicle in the field of view, are used to determine which portion of the frame is forwarded to a remote system. Again, compression as well as selective transmission of areas of interest rather than full frames can be used to minimise network usage.

Alternatively, the sensor may be configured to work with a variety of remote systems and adjust its transmission to suit each. Some smart sensors have built-in support for common industrial protocols such as Modbus as well as IoT protocols such as CoAP or MQTT. The sensor will use incoming requests to determine which remote node will receive the relevant data format. If network bandwidth requirements or sensor functionality determines that a single protocol should be used, gateways may be used to perform on-the-fly translation, forwarding Modbus packets, for example, to nearby PLCs and CoAP or MQTT packets to other systems that have subscribed to those data feeds.

A further advantage of the use of smart sensors is their ability to support secure communications, which can be combined with features that ease installation. A growing trend is for smart sensors to be shipped with their own digital certificates and private keys stored in protected memory. Some network protocols such as LoRaWAN have these facilities built into the system. When the sensor connects to the network, it can use those credentials to establish a secure link to a server using standard public-key cryptography techniques.

The use of stored credentials allows the sensor to identify legitimate servers and the server can itself verify that the sensor is valid. Only once that link is established is the sensor allowed full access to the network. Cloned or counterfeit devices will be identified and refused access. As cloud-based authentication systems make it possible to identify each device on the network individually based on its security credentials, the inclusion of these features can greatly ease installation. There is no need for an installer to program IDs and other information into the device as most of the necessary information will have been encoded during manufacture. If the sensor includes its own location hardware, using GPS or a similar system, even location may be determined by the module automatically. If not, either the installer or a remote operator can add the location and other metadata to the device and server databases once the sensor is up and running.

Once connected and validated, a secure smart sensor can add further protection to messages by encrypting packet payloads. In general, symmetric ciphers such as AES256 will be used for payload encryption because they have lower processing overhead than public-key systems. However, the architecture of the system and the performance of the sensor modules may favour the use of public-key encryption. Smart sensors may employ different ciphers for different users to ensure only data that should be available to a remote device is interpretable by that device. However, the system architecture may determine that this control over security is handled using edge gateways or cloud servers. There will be many possible combinations.

Read more on: Data manipulation in Smart Sensors Ecosystem

Smarter systems from sensor fusion

An important tenet of the IoT is that data from many different sensors is greater than the sum of the parts. A further implication of the extension of reach made possible by network connectivity is that many different sensor modalities can be combined to feed a data model or algorithm. The combination of different types of measurement makes it easier to determine if some inputs are suffering from errors caused by hardware failures or elements being blocked by dirt or grease. In turn, by rejecting errors in individual readings, the models will support better decision-making.

The application of sensor-fusion algorithms makes it possible to combine sensor readings in a coherent fashion. Some will use broadly compatible sensor formats. Sensor fusion is now commonly used in mobile devices. Sensor hubs built into mobile phones use the inputs from gyroscopes and accelerometers to improve the quality of applications such as gait analysis and navigation. The different sensors compensate for each other. The most important source of error in gyroscopes is that of drift. This drift can be compensated for by integrating data from accelerometers. In turn, the gyroscopes can help overcome the problem of sensor noise from which accelerometers suffer. The resulting outputs from the hub are more accurate representations of both linear movements and rotational changes such as roll, pitch and yaw.

The 360° views now provided by some advanced automotive systems are created by fusing the inputs from multiple cameras into one composite. Other systems use a variety of sensors to build a model for the system. For example, acoustic and vibration sensors may be coupled to help improve the fidelity of systems that look for damage to motors and other mechanical equipment. ToF cameras combined with temperature, CO2 and other environmental sensors may be used to help determine whether air-conditioning in a room or auditorium needs to be adjusted.

There are a number of effective techniques used to perform sensor fusion. One approach commonly used in motion-sensing systems is the Kalman filter. This is a filter that gives a higher weighting to readings that are thought to have low uncertainty. The filter state is represented as a set of matrices, which makes it possible to put readings from different types of sensors into a common coordinate model. The filter is divided into two parts, prediction and update. In the prediction phase, the filter calculates the next state of the system based on the previous state. In the update phase, freshly sampled values of the sensors are compared to their predicted values. The closer the input is to the prediction, the lower the error probability. If it is not a good match, the new reading for that sensor is given a lower weight.

Although it requires more processing time than the Kalman filter, the particle filter is tan effective option for situations where the data model is less linear than the systems for which Kalman filters are often employed. This type of filter combines input readings probabilistically using techniques such as Bayes' rule.

Probabilistic methods extend into techniques that call upon machine learning to perform more advanced types of sensor fusion. Machine learning is particularly suitable for systems that need to integrate multidimensional forms of data, such as 2D images, video and 3D point clouds from ToF cameras and LiDAR instruments. Deep-learning pipelines that combine multichannel convolutional layers with pooling provide a mechanism to train a model on various types of data in a consistent way.

An important aspect of sensor fusion is the use of pre-processing to ensure data elements are aligned. If some sensors are sending only changes at intermittent intervals while others are streaming, the receiving system will need to align and pad data values to ensure the model is updated with consistent values. For example, the model may need to be fed with repeated data values if a remote sensor has not sent indication of a change in state. Similarly, updates sent in groups will need to be aligned with the timestamps of other data streams to ensure that the samples are temporally consistent. Some of these functions may be taken care of by gateway modules or by end systems that have been programmed with the required understanding of the data they receive.

Read more on: Bluetooth 5 wireless technology for IoT smart sensors

element14
AN AVNET COMPANY

Conclusion

Electronics engineers are faced with many decisions, not just as sensing technology evolves but as the connectivity choices expand. The ability to detect changes around a system is greatly enhanced by the ability to hook into data from multiple sources on a network that covers the surrounding environment. Understanding how the sensors can interact with each other and their data processing and security capabilities will lead to smarter choices in designing with these products and integrating them to other devices.

For more whitepapers, please visit in.element14.com/technical-resources