The Latest SIM Developments —Evolving from Hardware to Software Profiles

2.1 SIM

2.1.1 The Advent of SIM Cards in Mobile Phone Systems

These days, we're all familiar with the concept of using SIM (Subscriber Identity Module) cards in our now quite affordable mobile phones. Anyone can swap or replace a SIM card with ease. We take them for granted now, yet they were not born at the same time as the mobile phone. Early mobile phones only supported "embedded" communications standards whereby the subscription parameters were hard-coded into device memory. The earliest analog standards like NMT-450 had no security features, which meant you could clone a mobile phone by copying the subscription parameters to another device. A well-known example of this in the wild from Japan is that of cloned pagers, which made it possible to broadcast messages to dozens of pagers using a single-device contract.

The first means of security came just a little bit later in the form of the SIS (Subscriber Identity Security) code, an 18-digit number unique to each device and hard-coded into the device's application processor. To prevent the same SIS code from being used on multiple devices, it was distributed evenly to carriers. The processor also stored a seven-digit RID code that subscribers send to the base station when registering with a mobile phone network. The SIS codes were distributed evenly among carriers so that no two devices could share the same SIS code. The processor also stored a 7-digit RID code which was transmitted to a base station when a subscriber registered to a mobile network.

The SIS processor would use a random number generated by the base station paired with a unique SIS response to generate the authorization key. Both the keys and numbers were relatively short but quite adequate by 1994 standards, but, as you can probably imagine, this system was later cracked. Three years later came the GSM (Global System for Mobile Communications) standard. This was quite similar to SIS, but it was more secure because it used a cryptographically stronger authorization system. Under communications standards from this point on, subscriber management on the device end thus became "detached." "Detached" meant that subscriber authorization all happened on an external processor integrated into a tiny computer completely separate from the mobile device. The resulting solution was the smart card-based SIM.

The arrival of SIM cards meant that (in theory) subscriptions were no longer device-dependent. This opened the door for device manufacturers to make mobile devices that would work on any carrier's system, facilitating mass production-driven cost reductions. It also meant that mobile users could change devices whenever and as often as they liked while keeping the same mobile identity.

SIM cards are basically based on ISO 7816 smart cards and are virtually the same as other contact IC (integrated circuit) cards like credit cards and cash cards. Indeed, the first SIM cards were the same size as credit cards, but as mobile phones became more advanced and the internal parts and components were increasingly miniaturized, SIM cards also became more compact.

The original full-size 1FF (1st Form Factor) SIM cards would no longer fit into mobile phones of the day, so a simple method of removing the unnecessary part of the card while retaining compatibility was developed. This was the mini-SIM, or 2FF (2nd Form Factor) SIM. Around the time these smaller SIM cards appeared, affordable mobile carriers in the form of MVNOs (Mobile Virtual Network Operators) were also starting to appear in Japan, thus leading to the roll out and widespread use of SIM cards.

SIM cards have continued to shrink with the arrival of the micro-SIM (3FF) and then the nano-SIM (4FF), yet their shape, the electrical contact configuration (pinout), and the features of the embedded IC chips has remained unchanged for around 30 years. To accommodate users who still cherish their old-school mobile phones, plastic SIM adapters are also now available. Even so, a lot of those old devices will not work with modern-day SIM cards even if the device will physically accept the SIM via such an adapter. This is because the earliest SIM cards operated on 5V, whereas the latest SIM cards run on

3V. That is, the processor voltage protection on 3V-only cards prevents them from working on older phones that can only accommodate 5V cards. Dual voltage SIM cards compatible with both 3V and 1.8V are also now becoming commonplace as the need for 1.8V cards rises amid the trend toward lower power consumption in mobile devices.

2.1.2 The Role and Real-world Status of SIM Cards

A SIM card is a small, highly secure, independent computer system detached from (independent of) the device on the mobile phone network system. It stores a dataset called the communications profile represented by an IMSI (International Mobile Subscriber Identity) and a 128-bit key called a Ki (key identifier). The SIM card connects to the mobile phone network system via base stations and is what enables safe and secure encrypted communications. The IMSI contains a Mobile Country Code (MCC) and a Mobile Network Code (MNC). MNCs are allocated to MNOs and full MVNOs.

IIJ Mobile obtained the MNC of 03 in 2018 when it became a full MVNO using the NTT Docomo network. It also obtained issuer number 03 at the same time. Physical cards, as defined by ISO 7816, basically have eight external contacts (pins). The pinout is shown below. The cards usually connect to the mobile device via six contacts: pins 1, 2, 3, 5, 6, and 7 (Figure 1).

An IC called a secure microcontroller is physically embedded within the SIM card's plastic. The IC consists of an MPU, ROM, RAM, and EEPROM—a fairly amazing and capable little system.

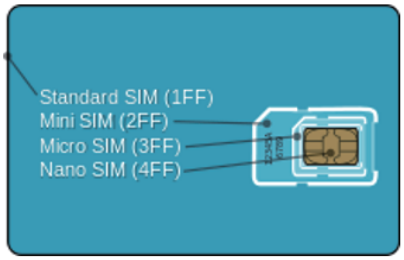
As they are computers, SIMs also have an OS. Many SIMs use an OS based on GlobalPlatform, the OS used in credit cards, which gives them an encrypted file system, the ability to run Java Applets, and both OS- and hardware-based tamper resistance. In terms of software, they have an encryption/decryption engine, as well as a communications profile—the dataset required to function as a SIM. Incidentally, credit card chips store a dataset called the financial profile, necessary for securing credit transactions.

All smart cards, including credit cards, have a unique 19-digit ID called the ICCID. The sequence of digits includes an industry identifier, country code, issuer number, and check digit. IIJ Mobile is able to issue SIM cards because it has obtained an issuer number.

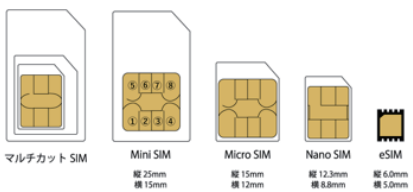
2.2 Toward a World Without Physical SIMs

Until a few years ago, users wishing to subscribe to IIJ's MVNO services online would first have to apply for a service contract, after which we would deliver a physical SIM card to their address, and then their service would only go live once they had inserted that SIM card into their device. The existence of the physical delivery step meant that users had to wait roughly a week before they could start using the service. SIM cards were, in effect, the physical keys used to gain access to mobile services.

A shift is rapidly underway, however, with the rise of eSIM services, which allow virtual SIM data to be downloaded to a device via the Internet, enabling instant access to mobile services. A similar trend is underway in the IoT world, too, whereby SIM data is embedded into cellular communication



Four example SIM card sizes that use the ISO/IEC 7816 interface.



ISO/IEC 7816-2 pinout

Pin #	Name	Description
1	VCC	+5 V or 3.3 V DC
2	Reset	Card Reset (Optional)
3	CLOCK	Card Clock
4	AS	Application Specific
5	GND	Ground
6	VPP	+21 V DC [Programming], or NC
7	I/O	In/Out [Data]
8	AS	Application Specific

Figure 1: The 8-pin Configuration of a SIM Card

modules at the factory before the modules are shipped to the device manufacturer, making it possible to use IoT services without a physical SIM card.

Throughout the evolution of wireless communications from 2G to 3G, 4G, and now 5G, the mobile industry has continued to use physical SIM cards, albeit with SIM card form factors growing ever smaller. Yet we are now approaching a turning point that will mark the end of an era and, with it, the end of the need for physical SIM cards to access mobile services. That era has spanned nearly 30 years and dates all the way back to the time of 2G (GSM) wireless communications. Let's go over the key changes below.

2.2.1 eSIM Support in PCs, Smartphones, Tablets, and Other Consumer Devices

In this article, eSIM refers to the mechanism set out in the Remote SIM Provisioning specification described in SGP.22, a standard put out by the GSMA (GSM Association), an industry association of mobile communications carriers, manufacturers, and the like. The adoption of eSIM makes it possible for users to gain instant access to mobile communications services by subscribing and then immediately downloading SIM data for that service to their mobile device.

Microsoft's Surface Pro LTE Advanced, released in 2017, was the first notebook PC to support eSIM, and all cellular-capable Surface models since then have been equipped with eSIM functionality. This has since fueled the uptake of eSIM functionality in cellular-capable Windows-based PCs from other manufacturers as well.

Support for eSIM has also become standard on many smartphones and tablets, including Apple iPhones and iPads since the 2018 release of the iPhone XS. Support for eSIM has also been available on Android devices since Google's 2018 release of its international-model Pixel 3, and the number of non-Google Android devices supporting eSIM has been increasing since then too. Hence, support for eSIM is becoming the norm in the world of consumer devices, with Apple leading the way.

In a further step in this direction, Apple released an eSIM-only (no physical SIM card slot) version of the iPhone 14 for the North American market in 2022, sending shockwaves around the industry. This is something we are likely to see more of in newly launched devices globally. We are moving headlong into a physical SIM card-free world for consumer devices.

With an eye on this trend, IIJ moved quickly to launch the SGP.22-compliant IIJmio Mobile Service Lite Start Plan (eSIM beta) on July 18, 2019, and has been progressively rolling out eSIM support on its services ever since.

2.2.2 SIMs on Cellular-capable IoT Devices

Unlike consumer handsets, cellular-capable IoT devices typically have a communication module and SIM capable of handling 4G (or the like) built in. IoT device end users often end up using the communication services provided by the IoT device manufacturer without realizing it, and thus do not necessarily enter into separate communications services contracts for their devices. In this scenario, the IoT device manufacturers procure physical SIMs from a mobile operator under contract in advance, and install them into the devices on the production line before shipping them.

Two types of requirements are increasingly coming to the fore in the world of IoT devices.

(1) The need for a physical alternative to SIM cards because either (i) the miniaturization of an IoT device has made it difficult to set aside space for a physical card or (ii) the device use environment is too harsh for an ordinary physical SIM card to withstand
(2) The desire to either (i) decide on which carrier to contract with after the device leaves the factory or (ii) change carriers depending on signal strength available at the device's eventual installed location

To address (1), MFF2, an IC chip form factor standard for smaller physical SIM cards developed by ETSI, a European standards organization, is already in use. In a further step forward, proprietary implementations that embed the SIM functionality into the communication module as software, such as SoftSIM, iSIM, and iUICC, are also coming into use.

Turning to (2), the GSMA released its SGP.02 standard for eSIMs for M2M (machine-to-machine) connections around 2013, before the release of SGP.22, but it has not gained a whole lot of traction in mass-market IoT devices because of the need to use services provided by specific carriers. This has led to people considering proprietary implementations that utilize the SGP.22-based eSIM framework, which is not tied to any specific carrier, and the new SGP.32 standard (which reuses aspects of SGP.22) for IoT released in 2023.

Where (1) is concerned, IIJ began providing MFF2 SIMs in 2019 as well as SoftSIMs combined with a specific communication module. In the case of (2), IIJ is engaged in a number of initiatives and studies, which we will discuss below.

2.2.3 Working Toward a World Without Physical SIMs

We are on the cusp of a physical SIM-free world, and the mobile services business based on the delivery of physical SIMs is approaching a major turning point. I think most people are aware that, from an end-user perspective, the spread of eSIM and other such technologies will make mobile services more convenient. For communications carriers, however, the impending disappearance of important elements like physical SIMs as the key to mobile services could be the writing on the wall if the carriers fail to adopt new technologies and allow themselves to fall behind the times. And as such, we at IIJ continue to carry out technical surveys, research, and development with a view to a physical SIM-free future. The next section focuses on our initiatives for IoT devices, which are a crucial area in particular.

2.3 IIJ Mobile's SIM Cards Applied solutions

Here, we go over a number of IIJ Mobile solutions made possible by rethinking the SIM computer system.

2.3.1 Multi-profile SIM

This solution makes it possible to selectively use multiple SIM cards without imposing a load on the device. Several logical SIM cards are set up on a single physical SIM card, and external instructions (APDU) are used to activate specific internal SIM cards. When multiple SIM sockets are available, this can be achieved by electronically switching access to the SIM socket. This is of course a DSSS (Dual SIM Single Standby) setup.

The idea is that, say, two half-thickness SIM cards are stacked and mounted into a SIM socket, and an external command is sent to switch access between the two SIM cards. Functionally, DSSS can work with even a single SIM socket (Figure 2).

2.3.2 SoftSIM

The required elements of a SIM include the MPU, ROM, RAM, I/O, OS, communications profile, encryption/decryption engine, and SIM communication protocol (APDU) implementation. IIJ Mobile's applied solution is SoftSIM.

This solution employs an eSIM-like approach. In simple terms, it uses computer virtualization technology to implement a virtual SIM (computer) in a secure area of the communication module, to which the separately managed communications profile is written OTA (over the air).

2.3.3 LPA-Bridge

It's a bit much to expect IoT devices to have rich UIs and multiple network interfaces like a smartphone, but devices equipped with the sort of sensors, LTE modems, and eSIM chips present in some smartphones can be considered IoT devices. The LPA (Local Profile Assistant, app used to manage eSIM profiles) on a smartphone is normally used to acquire, delete, and select profiles on its own internal eSIM chip. LPA-Bridge can be used to link with such IoT devices and switch the target of the LPA's operations from the phone's internal eSIM chip to the eSIM chip in the IoT device, so that the LPA can manage profiles on the IoT device's eSIM chip as if it were acting on the phone's internal eSIM chip.

This solution makes it possible to use consumer model eSIMs on IoT devices, something that was previously difficult to achieve via software without modifying the standard architecture.

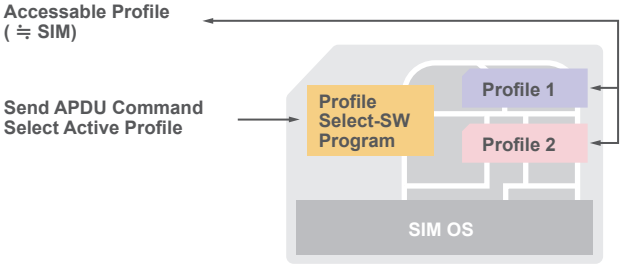


Figure 2: Selectively Using Multiple SIMs

2.4 Changes in eSIM technology Standards and IoT eSIMs

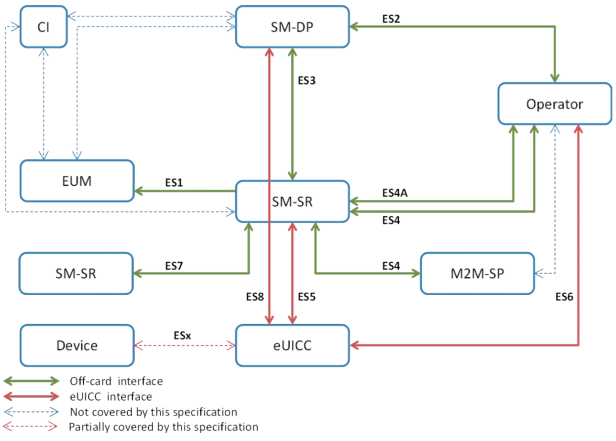
On May 26, 2023, the GSMA released SGP.32, a technical specification for eSIM for IoT devices. SGP.31/32 is the third eSIM specification, following the previously released SGP.01/02, which is for M2M devices, and SGP.21/22, which is for consumer devices. Below, we walk through the changes in eSIM standards leading up to SGP.32 and discuss some key features of the standard.

2.4.1 Road to IoT eSIM Standardization

As the name suggests (eSIM is short for embedded SIM), eSIM is intended to be an implementation of SIM that is embedded directly into a device's circuit board. Unlike physical SIM cards, eSIMs are difficult to replace once the device is manufactured, so the data that defines the SIM—the profile—is separated out from the hardware, and switching this profile effectively constitutes a SIM replacement. The mechanism for performing operations on the profile remotely is called RSP (Remote SIM Provisioning).

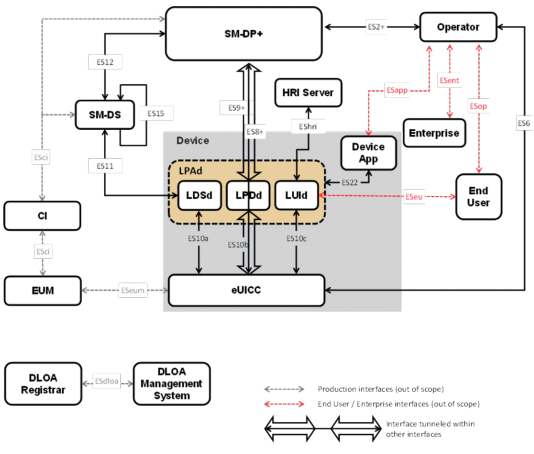
The first specification released was SGP.01/02 for M2M (machine-to-machine) devices (which we'll call M2M eSIM) (Figure 3). Perhaps because it was assumed that IoT devices would not have much complex functionality, most of the functions are implemented on the SIM, and the device interface is the same as with existing SIM standards. It is, however, a bulky system setup since the server (called the SM-SR) that communicates with the eSIM must be connected to the profile-providing server (called the SM-DP) provided by the carrier. And because SMS is used to trigger remote control, all of the profiles used need to have SMS functionality. SMS requires cellular communications capabilities, so a bootstrap profile is needed to ensure the device can connect to a cellular network in the location it will be deployed. Because the standard necessitates considerable cost outlays to build and operate the system overall, its use seems to be limited to high-priced automobiles and, in particular, the European auto industry, where the eCall system is popular and vehicles often travel across national borders. Speakers at international conferences have presented examples of independent carriers using the specification on smart meters in their country, but our impression is that they have really only used it to distribute their own company's profiles in the field and that they haven't been able to take full advantage of M2M eSIM.

The next specification released was SGP.21/22 for consumer devices directly operated by humans (which we'll call Consumer eSIM) (Figure 4). Because it is designed for devices intended to be directly operated by humans, an app (LPA) for facilitating this was introduced into the specification. Operations are performed via an LPA implemented on the device itself, so SMS (which was needed for remote operations) is no longer required under this standard, and IP is used uniformly for the data transfers used to acquire profiles. The standard also does away with the SM-SR, through which data transfers were relayed under M2M



Source: GSMA SGP.43 v4.3

Figure 3: The M2M eSIM Architecture



Source: GSMA SGP.22 v3.0

Figure 4: The Consumer eSIM Architecture

eSIM. Instead, the device communicates directly with the SM-DP (called the SM-DP+ in Consumer eSIM) provided by the carrier. This allowed for an open market not tied to any particular carriers, which would fuel widespread uptake of the standard. Indeed, once the Apple iPhone XS, for which there is a huge market, added official support for Consumer eSIM in 2018, use of the standard spread rapidly. Beyond Apple's iOS, Microsoft's Windows 10 and Google's Android 10 also came with LPA implementations. This meant that all the major OSes used on notebook PCs, smartphones, and tablets now had eSIM support, and this fostered an ecosystem in which the standard is available on many consumer devices. IIJ launched eSIM services on its full-MVNO platform ahead of its domestic peers in 2019.

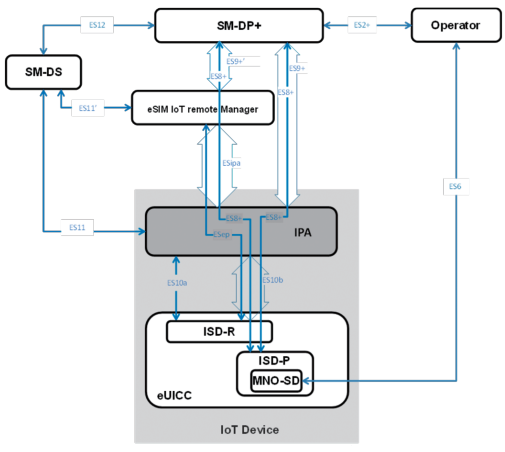
Consumer eSIM is designed for notebook PCs, smartphones, and tablets, but it also defines a mechanism for installing eSIMs on other devices via a smartphone or the like. This mechanism makes it possible to deploy Consumer eSIM on IoT devices that are not directly operated by humans. The GSMA standards, however, only lay out the architecture and do not define inter-device protocols, so at present, vendors implement their own protocols for this. The need to be linked to a smartphone or similar device has also meant that its deployment is limited to wearables like smart watches. It has not really gained much traction in the wider IoT device space. Against that backdrop and with the smartphone market becoming saturated, attention turned to IoT devices as the next target market. Ideally, M2M eSIM should have covered this area, but as discussed, the costs of deploying it can be prohibitive, and Consumer eSIM, meanwhile, requires proprietary protocol implementations to support the remote control features available in M2M eSIM. Hence, there was a need for an eSIM standard for IoT devices (which we'll call IoT eSIM). Development of GSMA standards is not open, and so vendor-hosted seminars and the like are the only way to keep track of what's happening, but from what we have heard, the GSMA began making some progress on the IoT eSIM front from around 2020. Ultimately, the architecture and system requirements (SGP.31) were released in April 2022, and the technical specification (SGP.32) was released in May 2023, thus standardizing the protocol (Figure 5). IoT devices based on this standard are expected to roll into the market ahead, paving the way for eSIM in the IoT device market, where the number of service connections is likely to far outstrip that in the market for consumer operated devices.

2.4.2 Features of the Standard

The IoT eSIM standard is designed to take advantage of the already expansive Consumer eSIM market. It uses the SM-DP+ from Consumer eSIM as the profile-providing server, and it follows Consumer eSIM with respect to the interface for communicating with the eSIM chip. It also adds the functionality necessary to enable remote operations. Because it reuses SM-DP+ from Consumer eSIM, no additional work to support it is required from the perspective of the carrier that provides the profiles.

It differs from Consumer eSIM in that the LPA functionality is divided between a server (called the eIM) and a device app (called the IPA), instead of being implemented entirely on the device. By providing an interface for the user (person operating the eSIM) on the eIM and having the eIM and IPA communicate with each other, it facilitates the remote operation of eSIMs on the device. Since IPA itself does not have a user interface, it has a smaller program footprint than LPA, making it easy to implement even on IoT devices with limited system resources.

The separation of functionality between the eIM and IPA appears to be a pretty flexible design for the purposes of supporting the vast variety of IoT devices out there. One major point is support for functionality called Indirect Profile Download, which makes it possible to communicate with the SM-DP+ via the eIM. The GSMA standard specifications define two methods for communicating between the IPA



Source: GSMA SGP.31 v1.1

Figure 5: The IoT eSIM Architecture

and the SM-DP+: Direct Profile Download (Figure 6) and Indirect Profile Download (Figure 7). With Direct Profile Download, the IPA communicates with the SM-DP+, so there is no need for SM-DP+ address resolution or HTTPS communications. With Indirect Profile Download, meanwhile, the eIM communicates with the SM-DP+, so the IPA itself does not need to perform address resolution or HTTPS communications. The IPA only needs to talk to the eIM.

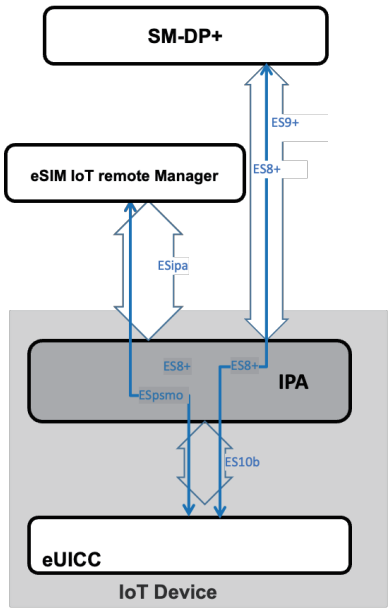
The GSMA standard specifications also define HTTPS and CoAP as the protocols for communication between the IPA and eIM, but any protocol is actually allowed (the appendix describes how to support LwM2M and MQTT), and support for non-IP communications is also considered. Indirect Profile Download makes it possible to use Consumer eSIM, which was designed for IP communications, on non-IP devices without any changes being needed to equipment on the carrier's end. It is possibly set up this way to also be consistent with the M2M eSIM architecture, which allows everything to be done via SMS.

2.4.3 Market Rollout

With Consumer eSIM now widespread, there has been talk in the past few years about IoT devices, particularly wearables, being the next target for eSIMs. While it may spur carriers to seek to increase service connection volumes, the fact that it obviates the need for physical SIM cards means that eSIM holds a lot of promise in the area of wearable devices, where physical space is scarce.

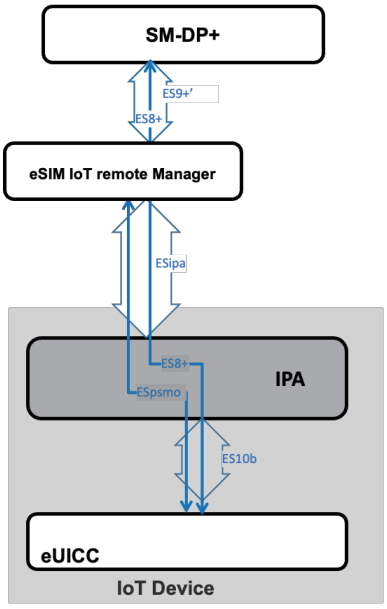
With Consumer eSIM, unlike M2M eSIM, you can simply select any communications carrier that offers Consumer eSIM, and this makes it relatively easy to use such devices even for small-scale rollouts. For manufacturers creating global models of their devices, it also has the benefit of allowing them to add local carrier profiles to the devices post manufacturing.

Challenges to widespread adoption remain, however. While you can use the same profiles as the Consumer eSIM, the issue of what to do about the bootstrap profile remains. The sort of devices that Consumer eSIM targets—laptops, smartphones, tablets, and the like—have non-cellular communications capabilities as well (e.g., Wi-Fi), so it was possible to ignore the bootstrap profile issue. Plus, smart watches and other such devices can communicate via smartphones, which also provides an avenue for installing profiles without a bootstrap profile. On IoT devices, meanwhile, things need to be implemented within resource constraints, so it may not be possible to include non-cellular communications capabilities, in which case a bootstrap



Source: GSMA SGP.31

Figure 6: Direct Profile Download



Source: GSMA SGP.31

Figure 7: Indirect Profile Download

profile is absolutely necessary in order to install the initial profile. Unlike with M2M eSIM, there is no clear bootstrap profile, so using a throwaway profile is fine, but unless the eSIM chip vendor or the IoT eSIM platform provider (not the communications carrier) provides an initial profile, IoT devices vendors may find it difficult to install one.

The implementation of IPA itself may also be a hindrance for IoT device vendors. It necessitates direct SIM access, so it is likely to be commonly implemented within a communication module rather than in a device app, but only a limited range of communication modules would be suitable for this. However, there is a method called IPAe, whereby the IPA functionality is implemented in the SIM, so if SIM card vendors provide IoT eSIM OSes that support this method, that may resolve the issue.

Competition with other systems is also an issue. Before IoT eSIM was released, Consumer eSIM also gained support for remote profile management (via the Remote Profile Manager, RPM) in version 3. According to the current specifications, the RPM functionality only supports the switching of installed profiles and not the adding of new ones. The GSMA is the standardization body in this case as well, so while all-out competition with IoT eSIM seems unlikely, developments in this area will bear close watching.

The IoT eSIM technical specification has only just been released, and test specifications required to validate interconnectivity (which we can assume will be released as SGP.33) are still in development, so a market rollout is still a little way ahead.

2.5 Conclusion

With a physical SIM-free world just around the corner, this article has looked at the current situation in this regard around smartphones, tablets, and the like as well as the situation around IoT devices. In particular, we discussed technical challenges that remain, along with the need for more testing and development, before IoT devices can go physical SIM-free, and we went over IIJ's efforts in this area.

Even in a physical SIM-free world, IIJ will continue to provide an environment for convenient mobile services while driving innovation that takes advantage of Internet technologies as it contributes toward the development of an increasingly networked society.

Daisuke Maruyama

Senior Engineer, MVNO Project Promotion Section, Technology Development Department, MVNO Division, IIJ
Mr. Maruyama had his start in the development of voice switches and later worked on the development of voice equipment for mobile phones before getting into mobile networks. He is primarily engaged in the study of technologies and development of services related to SIMs.

Munenori Ouchi

Senior Engineer, Mobile Platform Development Section, Technology Development Department, MVNO Division, IIJ
Mr. Ouchi investigates and engages in research on cutting-edge mobile technology, and develops services utilizing such technology.

Shigeyoshi Miura

Business Development Department, MVNO Division, IIJ
Mr. Miura has devoted 40 years to his work as an engineer. His career has spanned embedded hardware and software through to the design and development of DBMS applications and the design of system architectures. In recent years, his knowledge as an embedded systems engineer has proved quite useful, and he currently supports the development of mobile IoT devices and the application and use of SIMs.