Towards a 21st century telephone exchange at CERN

F Valentín, A Hesnaux, R Sierra and F Chapron
IT Department, CERN, Route de Meyrin 385, 1217 Meyrin, Switzerland

Abstract. The advent of mobile telephony and Voice over IP (VoIP) has significantly impacted the traditional telephone exchange industry—to such an extent that private branch exchanges are likely to disappear completely in the near future. For large organisations, such as CERN, it is important to be able to smooth this transition by implementing new multimedia platforms that can protect past investments and the flexibility needed to securely interconnect emerging VoIP solutions and forthcoming developments such as Voice over LTE (VoLTE). We present the results of ongoing studies and tests at CERN of the latest technologies in this area.

1. Introduction
CERN’s current fixed telephony infrastructure, designed and implemented more than 20 years ago, is unsuccessfully struggling to adapt to modern communication needs, notably the support of multimedia capabilities and subscriber mobility. The reason for this is that it was designed as a Time Division Multiplexing (TDM) network, a technology that has its origins with the beginning of telephony. Key technology developments of the past 20 years, notably voice over IP (VoIP) and the signaling protocol SIP (Session Initiation Protocol) cannot be adopted easily or at all and our infrastructure has reached a dead-end in its evolution. This paper sets out our plan for revitalising the telephony infrastructure at CERN through the adoption of 21st century technologies to replace those of the 19th century.

VoIP enables the transport of voice conversations, with the transmission of the original voice signal as a sequence of IP packets over today’s widespread IP networks. The formal definition of the signaling protocol SIP [1], marked a milestone in the maturity of VoIP developments, as it enabled the transport not only of voice but also of video, instant messaging, file transfer and any kind or combination of real-time multimedia.

SIP uses the underlying IP network to simplify enormously daily operation and subscriber management. With SIP, any user is able to connect a phone anywhere in the IP network and to change freely between different devices and even different technologies whilst keeping the same number—clear end-user advantages over the TDM model where a cabled phone line needs to be available. SIP also opens other communication channels for the corporate environment, where traditional Private Automatic Branch Exchanges (PABX) are being replaced progressively by Unified Communications (UC) solutions mimicking successful consumer-oriented applications as Skype. A second VoIP revolution is expected in the coming years with the worldwide rollout of fourth generation mobile networks, where the TDM domain will eventually disappear and all voice communications will be done using VoIP in the core networks where the IMS (IP Multimedia Subsystem) [2] built around the SIP protocol is a key enabler for VoLTE (Voice over LTE). A further VoIP traffic explosion will come from WebRTC (Real Time Communications on the Web) [3] that
provides an open framework for implementing native real-time capabilities in some of the most-used web browsers. WebRTC will expose the VoIP world to millions of new devices and brand new use cases, as many as can be fitted into the so-called Internet of Things (IoT).

2. CERN’s current telephony infrastructure
The central element of CERN’s current telephony infrastructure is a distributed PABX. In order to ensure service continuity and protect against power outages or IP network failures, four PABX nodes are installed across CERN’s campus, each equipped with an uninterruptible power supply (UPS). The TDM core handles switching and call routing functions; developments by the manufacturer over the last decade have enabled us to add IP links between the PABX nodes and support for IP phones. Although it is a robust and stable solution, the architecture has not kept pace with communication system developments and the monolithic rather than distributed architecture presents a challenge for integrating emerging technologies such as VoLTE, VoIP or WebRTC.

2.1. Services provided
CERN provides a variety of telephony services integrated into the PABX to its user population. Some examples are:

- End-user telephony, both with IP terminals and traditional analog lines.
- A manually-operated switchboard for outbound calling and information services.
- Emergency voice services from lifts and underground areas.
- SIP-based video conferencing (Vidyo) and unified communications (Lync).
- Three call centers, being two of them for infrastructure and user support and the third for the Fire Brigade and Safety services.
- Emergency digital radio communications using CERN’s own TETRA (Terrestrial Trunked Radio) network.
- Routing of outbound calls to the destination country on a least-cost basis
- Integration of mobile phones with the fixed phone environment, which, amongst other advantages, provides free mobile calls to CERN numbers and, for users connected to the local mobile network, outgoing calls at the CERN-negotiated least-cost rate.

2.2. User distribution
In order to accommodate the needs of the diverse user population at CERN different models of phones are supported (figure 1), of which the vast majority are analog phones, installed throughout all the facilities during the rollout two decades ago. These telephones and their evolved digital counterparts need a dedicated phone line, as explained in section 1. Given the associated infrastructure costs, especially operational costs, fixed line phones are deprecated and we have encouraged users to migrate firstly to IP phones and more recently to Lync; together these make up the second largest part of the user population, and the one with the highest expected growth for the coming years. The rest of the population consists of a mixture of emergency and lift phones with faxes, modems and virtual numbers not linked to a physical device for special use.

![Figure 1. Distribution of fixed telephony users](image-url)
3. A three-layer approach

To better understand the services supported by the PABX and its internal architecture it is useful to divide it in three different layers, as shown in figure 2.

Figure 2. Three-layered approach to the current fixed telephony network. The green arrows represent the sequence of internal actions taken internally when a PABX user calls an external number using the Least Cost Routing function.

The three layers match three levels of abstraction, with the devices and network connectivity in the bottom, which are used by the functions in the middle layer to establish the end-to-end conversations based on the internal databases (known as dial plans). The top layer uses the dial plan information along with the call details to provide billing and monitoring tools.

1. The lowest layer is the access layer, covering all the user phone terminals and phone lines, the emergency phones, a SIP gateway connecting to the Lync and Vidyo SIP domains together with connections to the external PSTN (Public Switched Telephone Network) operators and our mobile telephony operator. External connections are over 20 PRI (ISDN Primary Rate Interface) circuits each supporting up to 30 simultaneous voice channels.
   For this layer, the PABX is in charge of controlling the status of all the lines, providing the electrical power to the devices, keeping track of the operating status of the devices, managing the user profiles, applying special key configurations, controlling the congestion of all the external lines, as well as receiving the telephone hook signaling events on every line.

2. The middle layer is responsible for call management and handles the many different lower level devices simply as phone numbers that are part of one or several dial plans. The external dial plan is split into regions based on the price to call the destination so each subscriber is entitled to call to one or several regions. With the Least Cost Routing (LCR) function, the cost are minimized by finding the cheapest available operator to a certain destination.
   In this layer, the PABX knows the association between a phone number and a phone line, needs to manage the dialing rules for each of the dialing plans and needs to harmonize number formatting (short notation or long notation). Along with the LCR function for outbound calling, the PABX needs to handle the calling user rights and accept or reject calls based on them. Based on the congestion status of the lines in the lower layer, the PABX needs to find the cheapest feasible route to a destination.
3. Finally, the higher layer groups all the value added services like the call centers for emergency and infrastructure services, the switchboard calling and information service, the billing task that assigns a cost to all the external calls, the alarming system that sends SNMP traps whenever there is a hardware failure and an integrated SOAP (Simple Object Access Protocol) service that automates the creation of new users accounts.

It is in this layer that the PABX generates call tickets for all the calls in real time, and based on duration and external operator, assigns a cost. Additionally, the PABX maps all the physical events detected in any hardware element, and creates subscriber profiles mapped to phone lines in the layers below.

4. Project goals and target architecture

4.1. Replace the PABX by a software-based solution

The main limitations of the PABX systems are, firstly, an inherent monolithic design that is a blocking point for offering end users new convergent services (such as WebRTC or VoLTE), and secondly the high infrastructure, maintenance, hardware and licence costs.

After a survey of both commercial and open-source SIP solutions, we opted, given their equivalent level of maturity and to avoid the potential of vendor lock-in, to use open-source solutions at the heart of a state-of-the-art telephony network. With open-source software running on a combination of physical servers and OpenStack virtual machines through NFV (Network Function Virtualization) [5] and using SIP as a signaling protocol, our modular and scalable architecture is well suited to replace the current PABX and to provide a platform for the future addition of new services.

In this new architecture:

- The bottom layer remains unchanged, with the analog and emergency phones connected to a gateway that manages the line connection and signaling with the terminals and offers a SIP northbound interface to the middle layer. Also, the architecture will support a variety of softphones (software phones) in the near future and SIP-ready devices for conference rooms, wireless communications, intercom and machine-to-machine communications.

- A new entity, codenamed BRAINS, handles the call routing and subscriber management. BRAINS is a multimedia session router that takes routing decisions based on the originating and destination users solely and for every kind of user device, something that it is not possible in a PABX, where the devices not affiliated to the main domain are not able to access all the functionalities like call diversion or voice mailbox.

- The value added service will be adapted to interact with the new middle layer, in order to collect all the billing details from the gateways ruling each of the lower-layer SIP domains and monitor the status of the BRAINS entity.

The core idea behind this evolved architecture is to decouple the call routing and the other functions of the middle layer in figure 2 from the lower and higher layers. As can be seen in figure 3, BRAINS has been split into two functional elements, a front-end proxy and a routing engine.

- The front-end proxy serves as an interface to the gateways in order to control possible DDoS attacks by means such as white and black listing, checking of the originating and destination user, IP address filtering and dynamic promotion or demotion of trusted peers. Two open-source solutions are currently being studied for this element: Kamailio (formerly known as OpenSER) and OpenSIPS.

- The routing engine plays the key role in the call routing function. After receiving a SIP invite message validated by the front-end proxy, it will query its local user database to find where the destination peer can be found. If the user is reachable, the routing engine will reply with a 302 redirect message to indicate to the calling peer the next-hop IP address of the destination peer.
For the call routing engine a successful pilot test has been carried out with Asterisk, the most popular open-source IP PBX. A comparison of the relative advantages and disadvantages of Asterisk and its most popular fork project, FreeSWITCH, is planned.

4.2. SIP trunking with the external operators
While the above-described replacement takes place in the middle layer of our architecture, another goal of the project is to act in the interconnection with external operators by establishing SIP interconnections (SIP trunks). As described in section 2.3 these links are using still TDM technology, which limits each circuit to at most 30 simultaneous calls or 2 Mbps. These connections perform poorly when scaling the number of available channels, as more physical end-to-end circuits need to be established to operator premises. On the other hand, the available capacity of a SIP trunk is only limited by the IP link, quite higher than the 2Mbps of PRI links.

The BRAINS entity is not expected to do SIP header harmonization (even if SIP is an IETF standard, there are several interpretations about the usage of certain information elements of header that can lead to interoperability issues between SIP peers), so in order to ensure that no interoperability issues appear when connecting to the external operators and institutes, another functional element is needed. The functions of this element, called a Session Border Controller (SBC), include hiding the internal network topology from external operators, provision of access control mechanisms, SIP header normalization, media flow handling and transcoding (change of codec between two domains) [6].

The SBC will mediate between the BRAINS entity in the middle layer and the operator Customer Premises Equipment (CPE) as can be seen in figure 4.

4.3. Target architecture and future steps
Each of the BRAINS functional elements described in section 4.1 can be scaled independently. As can be seen in figure 5, a number, M, of front-end proxies will receive the SIP signalling messages from the lower-layer domains. All the front-end nodes in the cluster will share a DNS alias to make the redundancy transparent to the other peers. In turn, the front-end will use an internal look-up table to select one of the N routing engine nodes, implementing round-robin load-sharing. The routing engine nodes will have a local MariaDB database that is constantly synchronized with a central master database, where all the provisioning of new user profiles takes place.
Figure 5. Clusters of BRAINS functional elements. The right facing arrows mark the different call routing possibilities, while the left facing arrows indicate the synchronization from the master database.

Once the middle layer functionalities have been decoupled to the BRAINS entity and an SBC has been put in place to harmonize the communication with the external operators, the target architecture will resemble the schematic view depicted in figure 6.

Figure 6. Three-layered approach to the current fixed telephony network.

5. Summary and future work
The 20-year old TDM based PABX has a crucial role as the central element of the CERN’s telephony network but limits our ability to add support for new services and is costly to maintain and operate. A modern replacement is therefore needed and we have developed a replacement architecture that is based on open source software solutions and sufficiently robust and scalable to meet our large scale demands. A key element of the new architecture is the call routing and subscriber management functions for which we have developed an entity named BRAINS. BRAINS has been designed to be modular, scalable and future-proof, as one of the main goals of the network evolution is to accommodate new services to open new ways of communication between CERN users.

A two stage migration plan has been outlined, including the migration of call routing and subscriber management functions and the introduction of SIP connections to external providers with a
Session Border Controller to provide the necessary isolation. Although a solution for the replacement of value added services by SIP-based alternatives, especially for the call center and switchboard functionalities, has yet to be defined, our planned modernisation of the call routing, subscriber management and external connectivity will deliver a telephony infrastructure for CERN that is capable of addressing the challenges of the 21st century.

References
[1] Rosenberg J et al, June 2002, *SIP: Session Initiation Protocol*, Request For Comments 3261, Network Working Group, Internet Engineering Task Force
[2] *IP Multimedia Subsystem (IMS): Stage 2(Release 13)*, Technical Specification Group Services and System Aspects, 3rd Generation Partnership Project
[3] Ericsson A B et al, March 2015, *WebRTC 1.0: Real-time Communication Between Browsers*, W3C Editor's Draft
[4] March 1993, *I.431 : Primary rate user-network interface - Layer 1 specification*, International Telecommunication Union
[5] Chiosi M et al, October 2012, *Network Functions Virtualisation: An Introduction, Benefits, Enablers, Challenges & Call for Action*, ETSI White Papers
[6] Camarillo G et al, April 2010, *Requirements from Session Initiation Protocol (SIP) Session Border Control (SBC) Deployments*, Request For Comments 5853, Internet Engineering Task Force