Solution Brief

Logistics, Enterprise Safety, Manufacturing, Mining, Enterprise 5G, Industrial Automation, Warehouse

Sentra Edge-Deployed Push-to-Talk Solutions, Powered by Intel, Redefine Collaboration

Nybsys Sentra push-to-talk technology (PTT) bridges people, machines, AI and 5G, and scales to more than 8,000[i] PTT sessions on Intel® architecture

Push-to-talk (PTT) technology is a critical communication tool, even in an era of universal cellular connectivity. PTT offers unmatched instantaneous voice communication for a growing range of applications that demand quick response times and coordination. Traditionally, PTT has been used by first responders, but increasingly, it is being used by industrial organizations that need no-lag voice communications.

PTT capabilities are coming to 5G smart devices, transforming how PTT systems are used and making them more efficient and reliable. With a low latency of 1 ms, 5G networks do not add significant latency to PTT services, resulting in virtually no delay in communication, which is critical in high-stakes environments.

The vast capacity of 5G networks (1 million devices per square kilometer) allows for more devices to be connected simultaneously without compromising performance, facilitating seamless communication across large and dispersed teams. In industrial settings, PTT over 5G can support high-definition video calls, real-time data sharing, and Internet of Things (IoT) sensor integration.

In addition, Sentra PTT solutions provide enterprises with the unique capability to integrate generative AI-based intelligent virtual assistants with legacy communication devices. This can help reduce response times to operational events, which in turn can enhance safety and help mitigate risks. The system can instantly access and analyze vast amounts of data, deliver contextually relevant information, and offer actionable insights, greatly improving decision-making processes and boosting productivity.

Key beneficiaries of this integration are industries like manufacturing, warehousing, mining, and construction, where real-time information delivery can be used to monitor equipment health, manage workflows, and improve worker safety. With equipment and IoT sensors on the same 5G network, PTT can be used for simultaneous voice communications and real-time data streams from machinery, enabling faster decision-making and potentially reducing downtime.

Deploying private 5G networks in industrial environments is particularly promising. These networks offer secure, high-performance communication channels tailored to specific operational needs. This trend will likely accelerate as industries continue to digitize their operations and seek technologies that can enhance productivity and safety.

Nybsys is an Intel® Partner Alliance partner and Intel® Industry Solution Builders network builders community member. The company is an innovator of PTT solutions. The Nybsys Sentra PTT software enables enterprises to synchronize collaboration between personnel and machines, such as robots, sensors, and autonomous vehicles, to help improve productivity, safety and security.

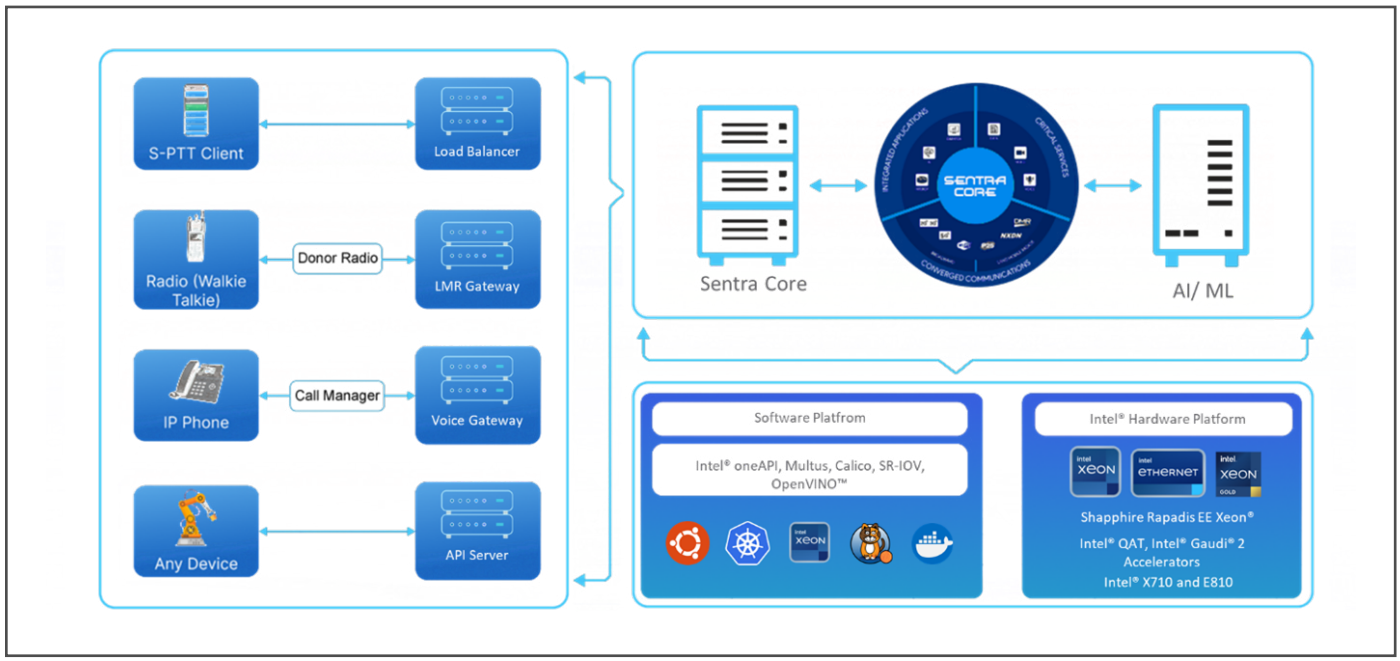

Figure 1. Sentra platform using Intel® Edge Technology.

The company has developed the software to run on Intel® architecture CPU-based servers and has used that hardware to test the scalability and capacity of the Sentra software.

Figure 1 illustrates the integration of Intel® technology within the Sentra solution. The Sentra Core, which manages integrated applications and critical services, is supported by both generative AI and machine learning systems and the underlying Intel® hardware and software platforms.

The hardware platform includes Intel® Xeon® processors and Intel® Ethernet Network Adapters X710 and E810, with single root I/O virtualizations (SR-IOV), which are essential for high-performance computing and network tasks. Intel® Gaudi® AI accelerators and Intel® Xeon® Scalable processors can provide reduced latency for generative AI applications using LLAMA 3.1[ii].

The software platform leverages the Intel® oneAPI Base Toolkit, Intel® Container Experience Kit, and SR-IOV, as well as open source Multus container network interface and Calico for network policy on Kubernetes to optimize the processing and network capabilities of the Sentra Core.

Generative AI with Retrieval-Augmented Generation (RAG) applications powered by Intel® OpenVINO™ Toolkit can enhance Sentra's capabilities by providing real-time video camera monitoring to alert users. The combination of these Intel technologies enables Sentra to deliver robust, scalable, and efficient communication solutions across its ecosystem.

Sentra Breaks Down PTT Silos

Nybsys Sentra is a PTT system that provides instant, secure, and context-rich collaboration. As shown in Figure 2, it supports nearly all network and device types used by first responder agencies or industrial enterprises.

Sentra software runs on servers designed for multi-access edge compute (MEC) workloads and are deployed at the edge to provide real-time communications with reduced latency.

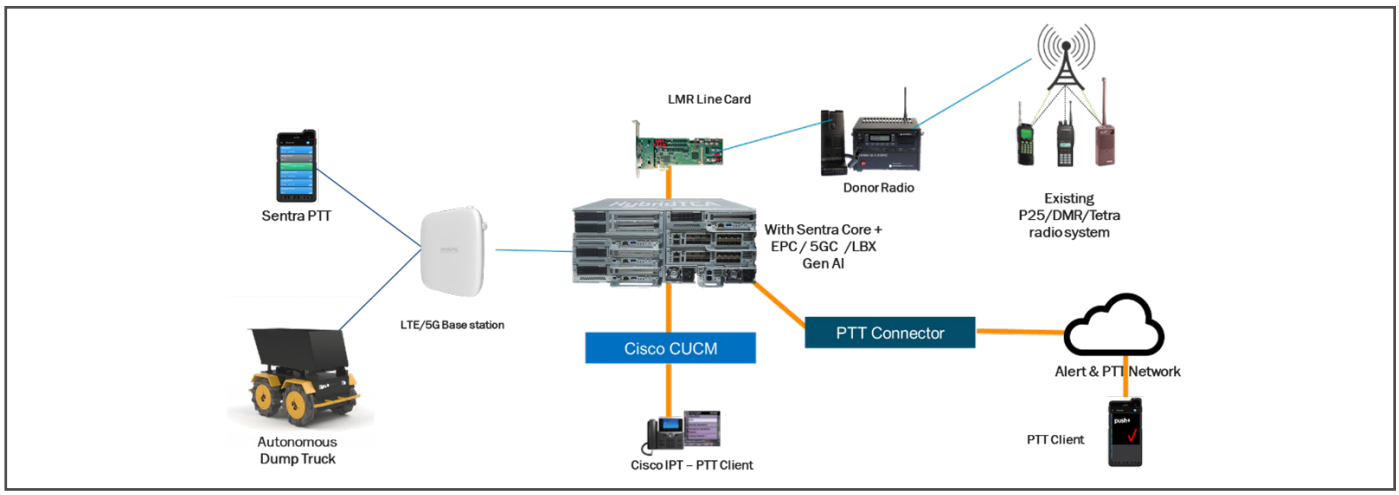

Figure 2. Bridging different networks to remove communication silos.

Figure 3. Sentra client and device support.

Figure 3 showcases the various clients and devices supported by the Sentra communication platform, which are all interconnected through the Sentra Core System. The Sentra IP Phone Client allows for traditional voice communication over IP networks, while the Sentra Two-Way Radio Client supports communication in environments requiring immediate and reliable interaction, such as in industrial operations.

The Sentra Mobile Client provides access to the Sentra system on-the-go, enabling users to stay connected through their smartphones and tablets. Additionally, Sentra AI Agent represents the integration of artificial intelligence within the system, offering automated support and enhancing communication efficiency. The Sentra Core System integrates these diverse clients and devices. It supports critical services like dispatch, data, and video, and is compatible with various communication standards, including 5G and Wi-Fi. This comprehensive platform ensures seamless, integrated communication across multiple devices and technologies.

The Sentra PTT Application Programming Interface (API) and Software Development Kit (SDK) are also available in an embedded version that can be deployed on IoT devices. This allows them to communicate alarms or changes in condition to the Sentra server, which relays them to PTT users for follow-up.

The 4th Gen Intel® Xeon® Scalable Processor-based server uses 100GbE Intel® Ethernet Network Adapter E810 and the Intel® Ethernet Network Adapter X710 for Ethernet connections to voice over internet protocol (VoIP) phones, such as the Cisco IP telephone (IPT), that can run the Sentra IP Phone PTT client. Also connected via Ethernet is a PTT connector that enables legacy PTT software from other vendors to connect to the network.

For dispatch applications, the Sentra Dispatch appliance is a full-featured, web-based radio dispatch that allows organizations to respond to incidents, emergencies, and other events. The system can track users' locations and status for more accurate responses. Dispatchers can assign users to PTT voice groups to support optimal collaboration.

The Sentra system can also connect legacy radios using a line card for land mobile radio (LMR) devices. This design consolidates islands of different generations of communications systems for interoperability, reduced costs, and improved reliability.

Sentra Generative AI Capabilities

Sentra Generative AI (GenAI) offers a force multiplier effect that can improve the effectiveness and efficiency of operations by providing insights and automating complex tasks. Sentra bridges traditional PTT users to the Gen AI platform, seamlessly integrating advanced AI capabilities into established communication systems (see Fig. 4).

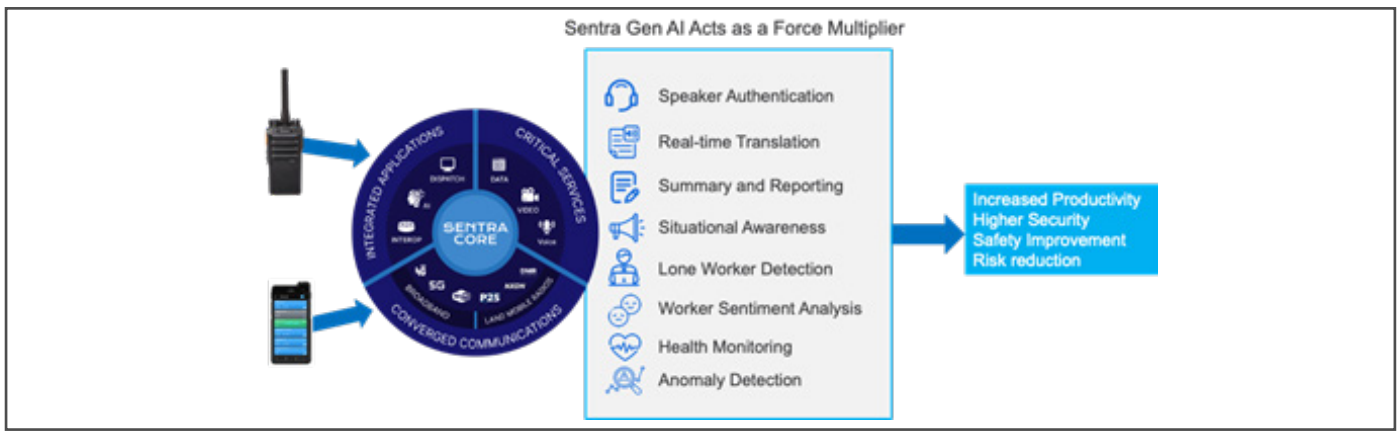

Figure 4. Sentra generative AI capabilities.

Sentra Generative AI Features

Sentra AI delivers substantial business value by integrating generative AI features that enhance operational efficiency, security, and decision-making.

Speaker authentication- provides secure access, ensuring that only authorized personnel engage with critical systems.

Real-time transcription and language translation- Sentra generative AI breaks down language barriers, facilitating global collaboration and communication across diverse teams and end users.

Summarization and reporting- Sentra summarizes the communications providing users with up-to-date communications.

Situational awareness- through continuous monitoring and analysis of environmental data, Sentra allows organizations to respond swiftly to changing conditions.

Lone worker detection- when a lone worker is detected, Sentra can send out verbal messages on a periodic basis listening for a response.

Sentra Bridges Traditional Generative AI Systems

Sentra seamlessly bridges the gap between legacy communication systems and generative AI-based technologies, effectively eliminating silos and creating a unified, cohesive system. This integration allows even traditional devices, such as two-way radios, to interact with generative AI voice assistants in real-time, ensuring all components work together harmoniously.

Figure 5 illustrates how Sentra acts as a bridge between traditional communication systems and modern generative AI-based applications. On the left side, the legacy communication system includes various devices such as smart phones with Sentra app, radio walkie-talkies, IP phones, SOS alert devices, and donor radios. These devices are connected to the Sentra Core through different gateways, like the app gateway, land mobile radio (LMR) gateway, voice integration gateway (VIG), and alert gateway.

The Sentra Core, shown in the center of Figure 5, serves as the central hub that integrates and manages these traditional communication systems while also connecting to advanced AI-driven applications on the right side. These AI applications include generative AI voice agents, RAG applications, and enterprise real-time data platforms.

These advanced applications interact with various modern technologies and services, such as robotics, access control systems, building management systems, video monitoring for situational awareness , intrusion detection, and weather feeds. Additionally, the diagram indicates that AI model training and threshold settings are managed through the Sentra Core to optimize these AI-driven applications. This means that enterprises can stay ahead of the curve, leveraging the best of both worlds: the reliability of legacy systems and the transformative power of AI.

The AI model can be trained and tuned with enterprise system feeds such as door access control, CCTV cameras, building management systems, intrusion detection systems, and robots. This integration allows a PTT client to receive AI assistance for enhanced situational awareness, enabling better collaborative incident response. Through the Sentra AI interface, even a traditional walkie-talkie can access the latest technology platform, transforming conventional communication devices into powerful tools for modern enterprise operations.

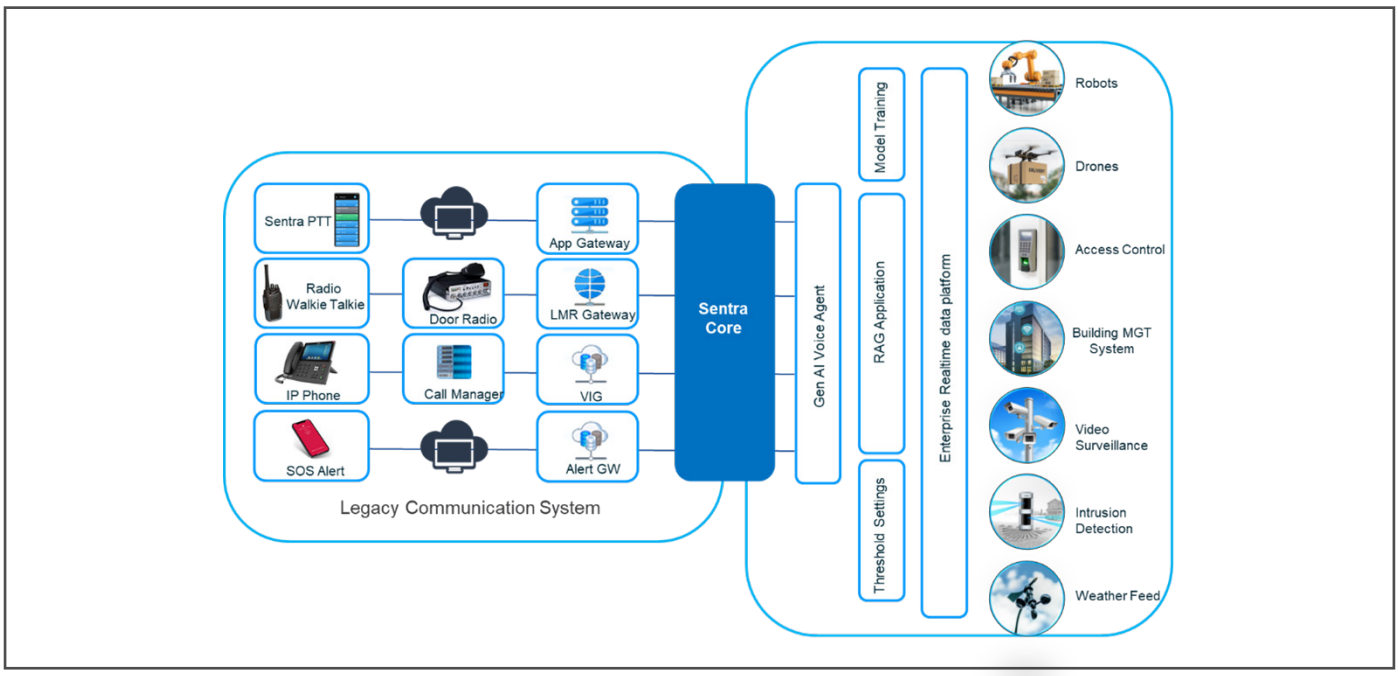

Figure 5. Legacy and generative AI system integration.

Scalability Test Set Up

The Sentra solution is optimized for Intel architecture processors and Nybsys undertook tests of the software on two servers powered by different CPUs to demonstrate the software's scalability. This scalability is an important consideration as PTT systems connect to increasing numbers of sensors, robots, and autonomous vehicles. Each of these connected devices and systems requires its own PTT stream for timely emergency or rapid response and improved safety.

Figure 6 shows the test set up. At the top are four high-performance PCs with Nybsys' own system load tester, each capable of generating 2,000 sessions simultaneously. This workload was sent to two systems under test (SUTs) running Sentra PTT software via a 10 GbE switch using SFP+ fiber cabling (see footnotes for SUT configurations).

The Config 1 SUT is a server running Intel® Xeon® Gold 6433N processor running a Kubernetes containerized version of Sentra. The Config 2 SUT is powered by Intel® Xeon® D-2166NT processor and is running the bare-metal version of Sentra. (See footnote for complete system configuration.)

The test compared the number of PTT sessions for a cloud-native Kubernetes-based MEC server using 4th Gen Intel® Xeon® Scalable Processors (Config 1 SUT) versus a bare metal server using an Intel® Xeon® D Processor. (Config 2 SUT. See footnote for complete server configuration.)

The 4th Gen Intel Xeon Scalable Processor family includes data-center class processors with up to 64 cores. These processors have the most built-in accelerators of any CPU on the market to help improve performance efficiency for emerging workloads, especially those powered by artificial intelligence (AI). In addition to performance improvements, 4th Gen Intel Xeon Scalable processors have advanced security technologies to help protect data in an ever-changing landscape of threats while unlocking new opportunities for business insights.

Intel Xeon D processors deliver workload-optimized performance in space- and power-constrained environments - from the data center to the intelligent edge. These innovative, system-on-chip processors support high-density, single-socket network, storage, and cloud edge computing solutions with a range of integrated security, network, and acceleration capabilities.

Sentra depends on high-speed Intel® Network Adapters to achieve low latency. Based on the network size, the company selected the 100GbE Intel® Ethernet Network Adapter E810 and the Intel® Ethernet Network Adapter X710.

Tests Show Scalable Performance

In this test environment, the 4th Gen Intel Xeon Scalable processor-based Config 1 SUT achieved up to 8,192 PTT sessions using a Kubernetes container configuration, which equals 256 sessions per core. The Intel Xeon D Processor Config 2 SUT recorded 1,382 sessions during the tests, with 115 sessions per core. In both scenarios, latency remained low at between 1 and 3 milliseconds[1].

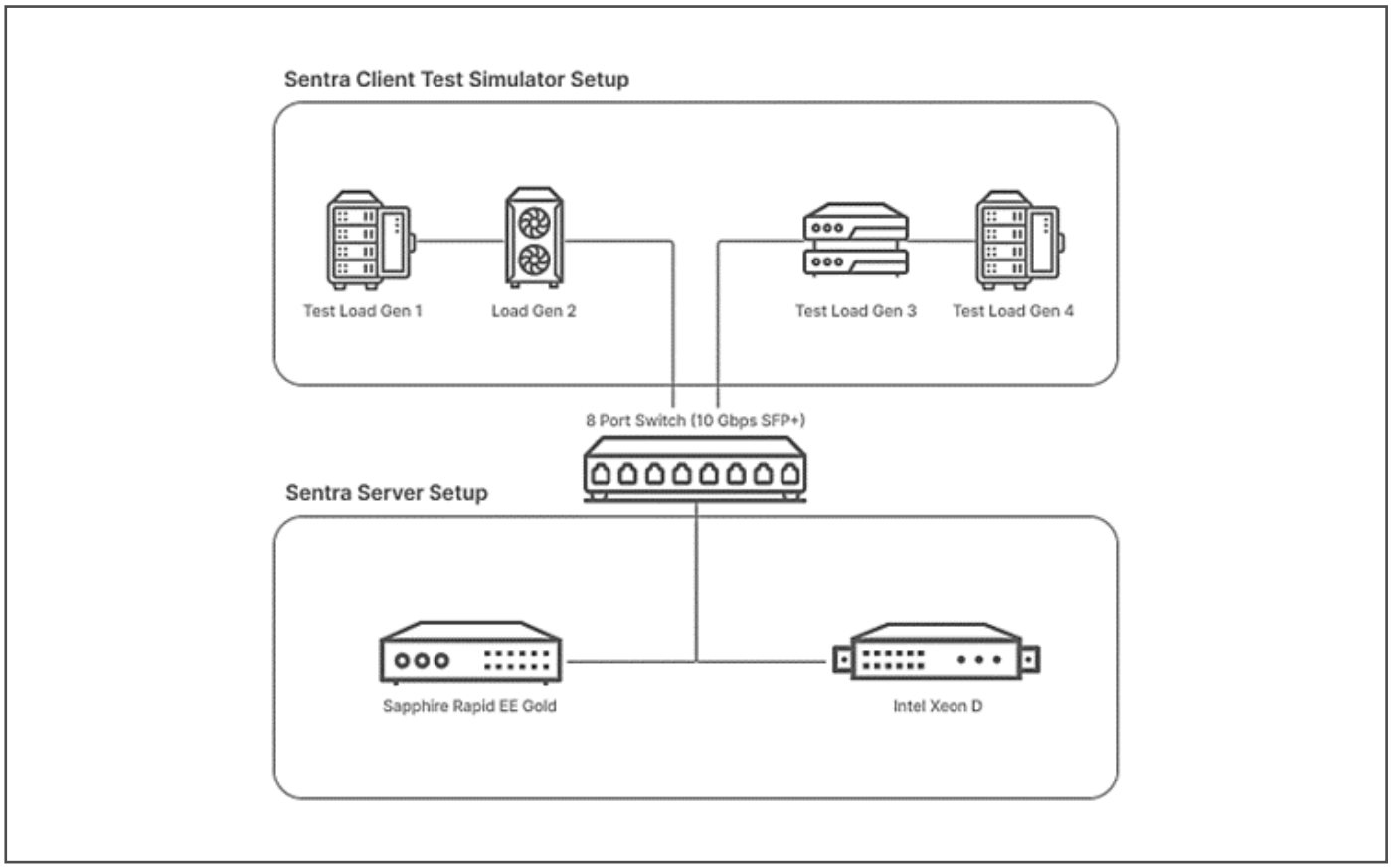

Figure 6. Sentra scalability test configuration.

Conclusion

Industrial enterprises and first responder agencies are turning to PTT and generative AI to improve operational safety and emergency collaboration among people and machines to improve team efficiency, reduce response time to system alerts and reduce occupational hazards. The Nybsys Sentra PTT system supports a diverse range of devices, including 5G-connected smart devices and IoT sensors, as well as legacy radios. This interoperability streamlines communications by replacing generations of radio systems that operate in technology silos.

With the growing use of IoT sensors, PTT systems need to be scalable to support fast growth in connected devices. The Sentra system with generative AI, running on 4th Gen Intel Xeon Scalable processors, achieved a capacity of more than 8,000 sessions. When the software ran on a lower-cost server with Intel Xeon D processors, it achieved over 1,300 sessions. In both tests, latency ranged from 1 ms to 3 ms. The tests show that customers can turn to Nybsys and Intel for the right-sized PTT system for their application. Nybsys can also integrate with existing customer systems and train the generative AI models adding additional Intel® compute resources at the edge.

Learn More

Nybsys Homepage

Sentra PTT

4th Gen Intel® Xeon® Scalable processors

Intel® Ethernet Network Adapter X710

100GbE Intel® Ethernet Network Adapter E810

Intel® oneAPI

Intel® Gaudi® AI Accelerators

[i]Config 1 SUT: 1-node, 1x Intel® Xeon® Gold 6433N processor (Lanner HCE 5040) with 32 cores and 64 threads. Total DDR4 Memory was 256 GB (8 slots/ 32GB/ 4800 MHz); HT on; Turbo off. OS: Ubuntu 22.04.4LTS; kernel 5.15.0-1054-realtime; 1 TB SSD; Benchmark/workload version: Sentra-K-RAN/01; Compiler: Intel® oneAPI DPC++/C++ Compiler 2024.0.2 GCC 11.4.0; Libraries: libnuma-dev, libhugetlbfs-dev, build-essential, cmake, meson Pkgconf Python3-pyelftools; Kubernetes v1.26.11 using Container Experience kit 24.01. Test by Nybsys on March 24, 2024.

Config 2 SUT: 1-node, 1x Intel® Xeon® D-2166NT processor (2.00GHz) (HPE P23465-B21 DI380 Gen10 4208 1P 32G Nc 8Sff) with 12 cores and 24 threads. Total DDR4 Memory was 32 GB (1 slots/ 32GB/ 2133 MHz); HT on; Turbo off. OS: Ubuntu 22.10; kernel 6.2.11-060211-generic; 1 TB SSD; Benchmark/workload version: Sentra Standalone on Docker; Compiler: GCC version 12.2.0; Libraries: Avahi, protocol buffers, ppc64el, riscv64, POSIX, libgcc-s1; Docker 23.0.1. Test by Nybsys on March 24, 2024.

[ii]Intel AI Solutions Boost LLMs: Unleashing the Power of Meta* Llama 3.1

[1]Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

Notices & Disclaimers

Performance varies by use, configuration, and other factors. Learn more on the Performance Index site.

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. No product or component can be absolutely secure.

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy. Intel technologies may require enabled hardware, software or service activation.

Your costs and results may vary.

Intel is committed to respecting human rights and avoiding causing or contributing to adverse impacts on human rights. See Intel's Global Human Rights Principles. Intel's products and software are intended only to be used in applications that do not cause or contribute to adverse impacts on human rights.

0924/CG/DJA/PDF Please Recycle 362213-001US