Sizing Environments for 2600Hz Global Infrastructure

We receive many questions from clients about how to go about predicting sizing environments for Global Infrastructure installments, and have written this guide to help in understanding this complicated endeavor. You pay 2600Hz to consult and spend a ton of time estimating and planning all the items listed below. The labor involved in that work is expensive, assumes you know what you're doing and understand the challenges. Unlike Private Cloud, you get to put your cluster anywhere you want in the world and you get full control over most aspects of it.

How do I get to 100,000 devices?

Context: Clients tend to tell us what they want to grow TO. They are concerned with making sure they're picking a long-term solution that they can keep and grow with. However, what they don't usually say up-front is that they don't want to pay for all of it at once. This makes things tricky because the cost to build an environment for 100,000 devices is inherently high, but the client subconsciously knows most of it will be idle as they get started. So, the client wants a path to grow to 100,000 devices, not an environment for 100,000 devices.

Global Infrastructure – it's limitless: Global Infrastructure is complicated because of the planning and coordination required, and often because of artificial barriers (such as a corporation's security department and audits). In general, most customers want to know, in advance, what to prepare for. We try to provide some rough guidelines but what most clients don't understand is that whatever they're going to prepare for depends directly on who they sell their product to and how the product is used. **Not all VoIP traffic is equal.**

The process of calculating how big to make your VoIP system is based on estimates and scientific calculations. How do you "size up" how big your cluster should be? The first question is what kind of traffic is it. The process of analyzing the traffic a client has is called "teletraffic engineering" - https://en.wikipedia.org/wiki/Teletraffic_engineering.

Teletraffic engineering is a bit like plumbing. Basically, you have a series of pipes, and you must make sure they can take as much water as is expected, but at PEAK demand. For example, you don't want to plan a building's sewage system on the assumption that only one person is going to flush the toilet at the same time. Rather, you might assume that if there are 20 units, they might all be showering and using the bathroom in the morning at the same time. Therefore, you must calculate how big your pipes must be to support that. You might also play some games like having areas where backups are temporarily okay. In VoIP, we're basically just glorified plumbers, working with packets instead of liquids.

What goes into the plumbing and related decisions? In VoIP, we typically look at a few things.
From a wholesale perspective (when speaking in thousands of users) the relevant points are usually:

- PEAK number of channels expected
- PEAK number of registered devices
- PEAK number of calls per second
- Total number of minutes per month

Typically, a traffic engineer calculates these for a large deployment (we can take on this role for each deal you set up or we can train someone to do this math, customer's choice, but it is important to have someone in charge of this).

Below we explain each item above in layman's terms. As an example, we'll assume an estimate for a 100,000-device customer.

PEAK number of channels expected

Channels are the most complicated to estimate, but also the most important for guaranteeing available resources (which impacts call quality). Channels represent audio (or video) being passed from point A to point B. A phone call typically involves a channel from the customer to the server, and another channel from the server to the carrier (two channels). Conference bridges, on the other hand, represent one channel per caller, all being mixed together. So, the number of channels can vary depending on application.

Without going into too much detail, in most environments, you can use these estimates when calculating your peak channels:

- **Residential:** If your environment is residential only, usually 1:10 (sometimes as low as 1:20) people are on the phone. So, if you expect 100,000 devices at residences, and if each "call" is two channels, that's 200,000 possible channels, but only 1:10 (or 1:20) expected to be in use. And only one channel goes to the carrier (the other is to the customer). That means, a good guess would be:
  - 10,000 - 20,000 possible channels (@82kbps or 1.56gbps IP bandwidth)
  - 5,000 - 10,000 possible SIP trunks / PSTN ports

- **Business:** If your environment is business PBXes, usually a 1:5 to 1:7 ratio applies during peak periods. If you expect 100,000 devices at businesses with normal PBX usage, and each "call" is two channels, that's 200,000 possible channels, but only 1:5 (or 1:7) expected to actually be in use. That would be:
  - 28,570 – 40,000 possible channels (@82kbps or 3.12gbps IP bandwidth)
  - 14,285 – 20,000 possible SIP trunks / PSTN ports
  - The above totals assume a light number of locations with call center and conference capabilities but not a focus on them

- **Call Center:** If your environment expects heavy / primarily call center usage, then the numbers are very different (hence why the expenses go up for call centers). You must plan for spikes, in many cases more channels in queue than devices to answer. In this scenario, we generally assume a 2:1 ratio (there might be twice as many people on hold as number of agents logged in), but you can sometimes get away with less. Depends on the call center and its PEAK volume expectations.

- **Conference Bridges:** If your environment expects heavy / primarily conference usage, then the numbers are very different, and most calls are inbound as well. You must plan for spikes and often dynamically spin up/down servers. This one is tougher and depends on how you market your product. Free Conference Calls inherently will have more traffic than paid ones, for example.

Assuming most of our clients are Business usage, then let's assume 1:5 to be conservative. So ~15,000 or 20,000 SIP trunks, and a good 4GB of bandwidth is probably safely sustained.

PEAK number of registered devices
This is the number of devices you plan to support. If you are assuming 1:1 device to user, then this is simply your user count. If you are assuming softphones and WebRTC clients, you need to worry only about how many you think will be online at any time. We've been using a 1:5 ratio to be safe for that estimate.
Without going into too much detail, in most environments, you can use these estimates when calculating your peak channels:

PEAK number of calls per second
This is more relevant for call centers and conference bridges, and represents the number of calls per second you think we'll receive in a sustained burst.

Total number of minute per month
Most residential users these days only use about 100-200 minutes per month. Most business users average 300-800 minutes per month. This is really for pricing exercises more than anything else in estimating your costs.

IF YOU DON'T KNOW THE ABOVE, THAT'S OK. Don't ask for numbers outside of number of devices – ask what service you are selling and number of devices only. That will tell you how to answer the above (or pass the info to us and we'll guesstimate).

We take the above numbers and then try to fit them into a platform/server/architecture. The channels make up the amount of volume (again, think plumbing) there is, and the other numbers make up the number of decisions we must make along the way (think valves, levers, etc. in plumbing).

Some rough budgetary estimates:

Starter Setup: 10K

Installation – Labor (assumes a 1 month install)
- Servers – 4 servers x $5,000 = $20,000.
- 1GBPS Core Routers, Firewalls, Switches = $8,000
  Total equipment per location: ~$28,000

Installation – Equipment, per location (2 locations minimum recommended)
- Network design / project management = $10,000
- Two rack/stack staff for two weeks = $15,000
- Sysadmin / setup / config = $25,000
- Misc travel expenses possible, depending on location

Ongoing Maintenance & Support

- Generally charged per-device and per-server
- Per-device is usually $3-5/device/month, depending on commit
- Per-server is usually $449/server/month, depending on commit
- Additional app fees may apply for call center/other services
- For general cost estimates, assume $3-5/device/month with a minimum commit of 5,000 devices
- May need to consider regular travel expenses if remote hands are not available at a facility

General Budgetary Estimate for 10K Starter Setup
Install: $56,000 equipment (2 locations) + $50,000 labor
Monthly Maintenance Commit for 10K Devices: $52,000/month

Expand to 25K Capacity:
Upgrade Equipment with New Servers, New Firewall/Routers (2 locations minimum recommended)

- Servers – 6 servers x $5,000 = $30,000
- 10GBPS Core Routers, Firewalls, Switches = $40,000

Total equipment per location: ~$70,000 upgrade fees

Installation – Labor (assumes a 1 month install)

- Network design / project management = $5,000
- Two rack/stack staff for a week = $7,500
- Sysadmin / setup / config = $5,000
- Misc travel expenses possible, depending on location

Total Labor: $17,500

General Budgetary Estimate for upgrade to 25K capacity:
Upgrade: $70,000 equipment
Monthly Maintenance Commit for 25K Devices: $125,000/month
Honestly past that it's sort of "as we go" pricing that averages out to $3-5/user plus small overhead in server fees that's spread across those users.