Canonical SIP Routing Platform (CSRP)

Product Overview & Capabilities Survey

Table of Contents

1.0 - Overview

1.0.1 - Abstract

The Canonical SIP Routing Platform (CSRP) is a turn-key SIP trunking service delivery platform powered by an advanced, high-performance least-cost routing (LCR), call accounting and rating engine. It was originally designed by wholesale VoIP operators desirous of a robust, high-throughput call processing core that does not compromise away sophisticated business logic capabilities.

However, it also extensively supports a retail-oriented SIP trunking feature set, including “light SBC” capabilities such as far-end NAT traversal, minute packages, surcharge assessment, and prepaid credit control. Outside of the wholesale ITSP space, CSRP has been used by:

- Hosted PBX companies;
- Short-duration (“dialler”) traffic movers;
- Calling card companies;
- Call notification broadcasters;
- Application service providers (e.g. hosted IVR).

CSRP is designed to run on the CentOS/RHEL distribution of Linux. It combines proprietary code developed by Evariste Systems with well-known open-source components, such as:

- Kamailio/sip-router SIP server and proxy;
- The PostgreSQL relational database backend;
- SIP Express Media Server (SEMS);
- Redis;
- Node.js.

Because these components are open-source, they increase robustness, ensure use of open Internet standards, maximise interoperability with third-party components, decrease initial and ongoing licensing costs, and lower overall operational expenses and personnel overhead by appealing to more common skills and more ubiquitous technologies.

CSRP is architected as an extensible framework that lends itself to dynamic evolution. This means its potential applications are not confined to the resale of per-minute PSTN origination and termination traffic, but extend more broadly IP multimedia communications.

1.0.2 - The argument for CSRP

The impetus for the engineering of CSRP was a perceived gap in technical and business-layer product capabilities for small to medium VoIP service providers specifically in the area of trunk routing and billing.

Most of the commercial options for VoIP ITSPs in this area aspire to be fully integrated softswitches. These products attempt to support some array of:

- Multi-tenant hosted PBX features, including extensive Class 5 and/or Centrex-style subscriber features.
- Application engines for products such as hosted IVR.
- Pre-paid applications such as calling card.
- Carrier routing, LCR, rating and billing.

In trying to be everything to everyone, these products often have extremely lopsided, inconsistent capabilities. One aspect of these platforms may be very sophisticated, while others may be very inflexible, limited, and poorly thought-out.

We have found that the carrier routing and LCR capabilities of many of these systems to be particularly lacking. The depth of sophistication required to accommodate real-world business needs is often absent in products that are stretched in many directions over a limited base of development resources.

Integrated softswitch vendors are under marketing pressure to claim well-rounded feature completeness, but underneath this superficial veneer of far-reaching capabilities in the "Class 5" space often lurks an inadequate "Class 4" implementation. Furthermore, the architectural requirements of a feature-rich subscriber-oriented softswitch that can be used to provide hosted PBX type services often conflict with those that maximise throughput and flexibility at the trunk level. The relatively heavy call processing workloads of a multi-tenant PBX platform are not necessarily appropriate for SIP trunking and PSTN interconnection.

Additionally, the ITSP in the wholesale or retail PSTN origination/termination business is faced with the unpleasant choice of either paying too much for an expensive infrastructure-based platform with limited intelligence (e.g. a classical SBC), or the technical burden and capital investment of "reinventing the wheel", in the form of a haphazardly-constructed in-house solution built from scratch.

CSRP squarely aims to address this deficiency. We believe that the market for multi-tenant hosted PBX platforms and application-oriented softswitch assemblies is adequately served.

CSRP does not try to be everything to everyone. Its technical orientation, architecture and feature set is **geared for retail SIP trunking and wholesale ITSP operations.** Our target customer is looking for an affordable platform that **does one thing and does it well**.

CSRP can be--and, with some frequency, has been--combined with an incumbent "Class 5"-style softswitch and/or feature platform. Such a configuration allows the customer to benefit from the most optimal intersection of specialised capabilities; the "Class 5" platform excels at what it does, while CSRP delivers high-end capabilities for the trunk routing and rating portion of the overall service offering.

1.0.3 - Technical architecture

Fundamentally, CSRP consists of a Kamailio SIP proxy backed by a PostgreSQL database instance. Much of the business logic involved in call processing is contained in PostgreSQL stored procedures that are called by Kamailio. Some of these stored procedures are invoked directly from Kamailio route script, while others are called by triggers that fire in response to row-level changes to the underlying database made by Kamailio.

This architecture minimises the costly overhead of unnecessary data interchange between the database and the proxy, restricting it to end-results that are consumed by the proxy to effect call processing. The stored procedure execution environment has several additional benefits:

- It is inherently transactional, which minimises the risk of data corruption. If a procedure call fails with a fatal error, all changes made in the course of the transaction are automatically rolled back, which prevents the data set inconsistencies that would arise if the underlying queries were issued discretely by the proxy and only some of them failed.

- It allows many updates to the business layer without any need to restart the proxy.

As with all architectural decisions, this approach entails certain technical compromises and performance trade-offs. One is that the preponderance of the workload is database-bound. As a result, storage I/O performance and memory available for database-tier caching greatly influences performance in the upper range of the platform's throughput capabilities. Where appropriate, in-memory, runtime data structures are utilised to ensure high performance.

At the same time, a closely database-bound call processing workload confers many advantages in that it leverages sophisticated database-level optimisations and assurances for data integrity and internal consistency--advantages whose reproduction outside of a database often results in significantly inferior, less reliable imitations of the thing they are trying to avoid.

1.0.4 - Software elements

A minimal CSRP deployment consists of the following mandatory software elements:

- A Kamailio SIP proxy;
- A PostgreSQL database instance;
- Node.js API middleware;
- Nginx web server, to serve web-based provisioning and administration interface, and to route the API calls it utilises;
- A Redis cache server backing the proxy and the API middleware.

Optionally, one or more outboard RTP relays may also be utilised to relay media through CSRP. CSRP supports two different kinds of media relays based on throughput needs. For more information, please see the product FAQ.

1.0.5 - Redundancy

CSRP can be deployed in redundant configurations of varying degrees to achieve a yield at the optimal intersection of risk management and budgetary constraints desired by the customer.

For more information, please see the product FAQ.

2.0 - Capabilities survey

2.0.1 - Call accounting

The CSRP SIP proxy generates elaborate call detail records (CDRs) and posts them directly to the database in real time. Intermediate export facilities such as RADIUS are not used, both for performance and functionality reasons. However, like RADIUS-formatted call accounting records, CSRP CDRs are exported as one data row per call.

To ensure maximum throughput performance, CDRs are written--and any associated database trigger cascades--in an asynchronous fashion. This means that SIP message processing threads in the proxy do not block on call accounting operations, leaving them free to process more calls.

2.0.2 - Call rating

CDRs are rated in real time by database triggers.

For both origination and termination rating, separate rating is supported with respect to the rates assessed the end-customer and computed vendor costs. In this way, CSRP simultaneously accounts for minutes billable to the end-customer and vendor costs incurred.

This permits the user a high degree of immediate, real-time visibility into profit margins, as well as a basis for vendor bill reconciliation.

The vendor costs and retail rates are entered separately into the database and are not computationally linked. Thus, the user is free to assign any rates to the end-customer for both inbound and outbound calls, and this computation is completely irrespective of the underlying vendor costs. The effect is that CSRP does not force the user into a naive pricing model where retail rates must be a simple linear derivative (such as a percentage markup, for instance) of underlying vendor costs. The user is free to adopt any desired pricing model, separately configurable for each customer.

2.0.3 - Concurrent call limits

Enforcement of concurrent call limits for both origination and termination is supported. Separate limits can be configured for both end-customer endpoints and vendor gateways.

2.0.4 - Dialed digit and ANI translation

Translation operations (stripping digits/characters, prepending an alphanumeric prefix) are supported for both origination and termination endpoints, and can be applied to both dialed digits (DNIS) and ANI/caller ID.

2.0.5 - Performance metric collection

CSRP collects--on a per-call level--and exposes aggregate reports of various metrics related to call processing and routing health. These include, among others:

- **Post-dial delay** (PDD) - The temporal gap between relay of initial SIP INVITE message and some form of successful provisional progress and/or ringing feedback (non-100 1xx message).
- **Answer-seizure ratio** (ASR) - The ratio of calls attempted to calls answered.
- **Average call duration** (ACD) - Average duration of answered calls.
- **Profitability** - Average and cumulative profit per customer, per vendor, per call, per destination, etc.
- **Costs** - Cost summary per destination, and/or per vendor.

2.0.6 - Static IP and SIP registration routing

CSRP includes an integrated SIP registrar. It fully supports routing and authentication to and from end-customer endpoints in one of two ways:

- **Trusted IP source/destination** - Commonly necessary for interoperation with intra-industrial VoIP equipment such as softswitches and media gateways, as well as some PBXs and ATAs. This equipment does not commonly support the SIP REGISTER method or digest challenge authentication.

- **SIP registration/digest challenge authentication** - Most commonly used with end-user PBXs in SIP trunking scenarios over the public Internet. This often--though not necessarily--is used in tandem with username and password digest challenge authentication/authorisation on outbound calls.

2.0.7 - Credit control

2.0.7.1 - Pre-paid credit

CSRP supports the optional provisioning of pre-paid credit to a customer endpoint. The balance of this credit is deducted after a call completes and rating data for it is posted. The minimum balance threshold that must be met for calls to be completed is also configurable, and can be set to a positive or negative value.

CSRP is capable of automatically generating e-mail notifications to customers and/or vendors when the prepaid balance drops below the minimum threshold, thus encouraging them to add more credit to their account.

As well, CSRP supports the assignment of credit control operations for a given customer billing group to a "target" billing group, permitting flexible reseller arrangements or master-child trunk relationships.

2.0.8 - Termination capabilities

2.0.8.1 - Least cost routing (LCR)

CSRP contains a very high-performance, database-backed least cost routing (LCR) engine that can search hundreds of millions of variable-length prefix routes in milliseconds. This component has been tuned to meet the peak requirements of users with extremely high-frequency outbound dial traffic.

The LCR process returns a pre-loaded route set to the CSRP proxy, allowing rapid fail-over through up to 30 termination vendors using serial forking. Failover can occur in the event that attempts to contact a vendor gateway time out, or if a vendor gateway explicitly rejects a call.

Round-robin distribution of calls among a termination vendor's gateways is also supported.

2.0.8.2 - Inter/intra-jurisdictional routing

In many regulated telecommunications jurisdictions, termination costs fall into a dual pricing scheme depending on whether the origin of the call--as determined by the ANI--is within that same jurisdiction or outside of it.

For example, in the United States, where CSRP was developed, termination vendors commonly offer two pricing tiers depending on whether a call is intra-state (originates in an

NPA/area code belonging to the same state as the NPA/area code of the destination) or inter-state.

CSRP accommodates this scenario in a flexible way with user-defined regional code groups. Thus, the CSRP LCR engine takes inter/intra-regional pricing possibilities into account when ordering routes by ascending cost preference.

2.0.8.3 - Vendor termination route groups

CSRP does not have a "global" termination route table. Rather, the set of vendors from which the LCR process picks is known as a "vendor termination group". It is possible to define an arbitrary number of these groups and assign different groups to different customers.

Thus, it is possible for the CSRP user to sell only routes from certain vendors to a given customer. This capability is frequently used to create differentiated "premium" termination offerings.

2.0.8.4 - Local number portability (LNP) dip over SIP support

(North America/NANPA only.)

In the United States, where CSRP was developed, actual termination costs faced by carriers vary with the carrier that is the final recipient of the call, not dialed destination. It is increasingly common for termination vendors to shift the business risk arising from the possibility that a number has been ported (to a different carrier than the one implied by the dialed number) onto the end-customer by billing based on terminating OCN (Operating Carrier Number) and LATA (Local Access and Transport Area), not dialed number.

To properly mitigate this risk, the end-customer must perform an LRN (Local Routing Number) database query (known as an "LNP dip") to determine the actual terminating OCN/LATA of the dialed number.

CSRP supports a common means of exposing an LRN query interface via SIP using redirect replies, as described more abstractly in RFC 4694. In this scenario, CSRP will query an LNP dip provider and extract the LRN from a redirect reply, if available. It will then perform normal LCR routing on the LRN, rather than the dialed digits. Storage and caching of the LRN is also supported.

Finally, failover routing among an arbitrary number of LNP gateways is supported, allowing the use of primary and secondary (and beyond) LNP query vendors.

2.0.8.5 - Profit margin enforcement/stop-loss

Optionally, the CSRP LCR process can be instructed not to choose vendor routes whose per-minute cost exceeds the retail termination rate that is assessed to the customer for the given call. The cost differential threshold is configurable and can be a positive or negative value, allowing the CSRP user to permit a limited degree of loss as well.

In addition, the LCR process can be instructed to not return vendor routes whose minimum billing duration exceeds the minimum billing duration being assessed to the customer for the given call.

2.0.8.6 - Termination rate plans

Termination prices, together with minimum durations, billing increments and prefix specificities, are grouped into termination rate plans. Different termination rate plans can be assigned to customers, providing maximal pricing flexibility.

Termination plans can also inherit from other termination rate plans, reducing the need to maintain redundant route entries.

2.0.8.7 - Maximum cost thresholds

Per-customer maximum costs can be defined for both retail rates and vendor routes. That is to say, this feature allows one to provide to a caller only retail routes that cost a certain amount or less, or to exclude from LCR choices all vendor routes that exceed an absolute maximum per-minute cost.

2.0.8.8 - “Tech prefix” support

A single customer billing group (BG), identified by an IP address or registrant, can select from multiple termination rate plans and vendor termination group combinations using a prefix map. This is commonly known as a “tech prefix”. The effect is to permit one customer endpoint to choose from multiple termination products.

2.0.8.9 - Harnessing of customer LNP (NPDI) parameters

Some wholesale termination customers can perform their own LNP dips prior to routing calls to CSRP, and do not wish to be double-charged for another dip. CSRP supports extraction and harnessing of customer-provided LNP routing data.

2.0.8.10 - Termination block list

Destinations in termination rate plans can be explicitly blocked, permitting the creation of a blacklist of any desired prefix granularity.

2.0.8.11 - Termination ANI blacklist

Offending customer ANIs can be blocked from passing outbound calls, by a prefix of any length.

2.0.8.12 - Post-dial delay-based routing failover

Outbound routing can be configured to fail over to the next available termination vendor route if a vendor gateway does not provide progress indication feedback or answer within a certain period of time. This allows the control of post-dial delay experienced by the caller as result of vendor call completion delays.

2.0.9 - Origination capabilities

2.0.9.1 - Longest-prefix match routing

Much as termination, origination call processing utilises two independent routing tables with respect to vendor and customer routes. Fast longest-prefix matching is applied to both, allowing different route granularity between the vendor and the customer routing tables.

Thus, a CSRP user with a large, contiguous number block that falls on a decimal boundary on the vendor side can simply enter that block as one route, but break it out into longer (more specific) routes on the customer side. Because longest-prefix matching is always used, more specific exceptions to broader routes always have precedence.

This approach allows routing tables and DID inventories to stay as small and manageable as possible.

2.0.9.2 - Failover among customer gateways

Customer SIP destinations for origination are processed in ascending order of priority, and can consist of any combination of static IP/port destinations and registration AORs

(Addresses of Record). This allows the CSRP user to offer customers maximal redundancy through routing to secondary (or greater) PBXs or gateways.

2.0.9.3 - Flexible DNIS override in Contact

Every route to a customer registered AOR (Address of Record) contains a setting indicating whether to adhere to the Contact bound to that AOR strictly, or to pass in the dialed digits in the user part of the request URI. This allows PBXs that route on DNIS to receive dialed digits regardless of what Contact they actually registered with their binding, while preserving interoperability with other devices that expect to receive precisely the request URI that they registered with.

2.0.9.4 - Voicemail backup

In the event that no customer endpoints can be reached on an inbound call, CSRP can, as a last resort, route calls to its internal voicemail-to-email application. This behaviour can be configured on a per-route basis. With this feature, the customer can receive voicemails even if all of their endpoints are impaired.

2.0.9.5 - PSTN call forwarding support

Inbound numbers can be forwarded, through the B2BUA component, to a PSTN destination. Both legs will be appropriately billed to the customer--the origination leg at the appropriate origination rate, and the termination leg at the applicable termination rate.

2.0.9.6 - Inbound round-robin distribution

Inbound calls can be distributed in a round-robin fashion, with failover, among all of a customer's destination gateways. This is ideal for routing to a cluster of application servers providing services such as hosted IVR or calling card.

2.0.10 - Automatic CDR export

A facility for the high-speed export of call detail records into comma-separated values (CSV) format is provided.

This is ideal for customers who do not wish to use CSRP's built-in call rating, but instead would like to import CDRs into a post-rating system for further processing and billing.

2.0.11 - Rate import

The problem of importing bulk quantities of vendor and customer rates from comma-separated values (CSV) input format is provided for by a high-speed data import processor, ensuring minimal disruption while large rate tables are reloaded.

2.0.12 - Call setup rate control

CSRP can limit call setups per second (CPS) by customer endpoint or vendor endpoint. The call setup rate for each endpoint is sampled at a configurable interval (usually 3 or 5 seconds) and calls that exceed the threshold are rejected.

2.0.13 - Minute package support

CSRP supports packages of minutes, including overage billing, for more flexible retail offerings.

Package credit is applied to both origination and termination calls. Not all termination calls need be covered by a package; a list of prefixes for each package is used to make that determination.

In the event that the user exhausts their package credit in the middle of a call, the overage portion is billed at a per-package overage rate.

2.0.14 - Taxation surcharge support, “USF-ready”

CSRP supports calculation of fixed and variable surcharges using discrete surcharge profiles.

US “USF” (Universal Service Fund) taxation support (percentage of interstate revenue) is built-in. A surcharge profile can be customised for other jurisdictions with different taxation rules.

This mechanism also allows the implementation of a conditional surcharge for short-duration calls, which is frequently used for revenue management associated with “dialer” traffic.

2.0.15 - Cost-based fraud control

A key fraud protection feature of CSRP is to limit a customer to a configurable amount of simultaneous calls whose per-minute cost is equal to or higher than a certain configurable amount.

If this limit is exceeded, all existing calls will be deactivated, and the customer can be switched to a “safe” termination rate plan that excludes high-cost destinations. E-mail notifications of this event will also be generated.

2.0.16 - Topology hiding

CSRP supports SIP-level topology hiding. When combined with RTP relay, full topology hiding can be provided.

For more information on how the exact mechanism, please see the product FAQ.

2.0.17 - CDR archival on Amazon S3

CSRP supports periodic upload of CDR exports to Amazon S3 cloud-based block storage.

2.0.18 - CDR upload to Amazon RedShift

CSRP supports periodic synchronisation of CDRs to Amazon RedShift, a cloud-based column-oriented database service. This is ideal for large analytical tasks and expensive reporting operations.

2.0.19 - Signalling-level dead peer detection/keepalive

CSRP can issue a periodic SIP OPTIONS ping to both parties in an established dialog to check if they are still responsive. The call will be terminated if one or both endpoints do not respond. This helps eliminate “stuck” or “dead” calls.