Evariste Systems

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 is intended for use principally by wholesale VoIP operators that require large-scale call processing with sophisticated business-layer capabilities.

CSRP has been deployed by numerous pure-SIP ITSPs (Internet Telephony Service Providers) and carriers as a core routing and switching platform to deliver PSTN termination and origination.

CSRP is designed to run on 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);

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 an off-the-shelf, turn-key product. However, its architecture as an extensible framework accommodates customisation. This allows it to be applied to a range of next-generation product and service concepts.

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 the platform may be very sophisticated, while another may be very inflexible and limited.

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 often conflict with those that maximise throughput and flexibility at the trunk level.

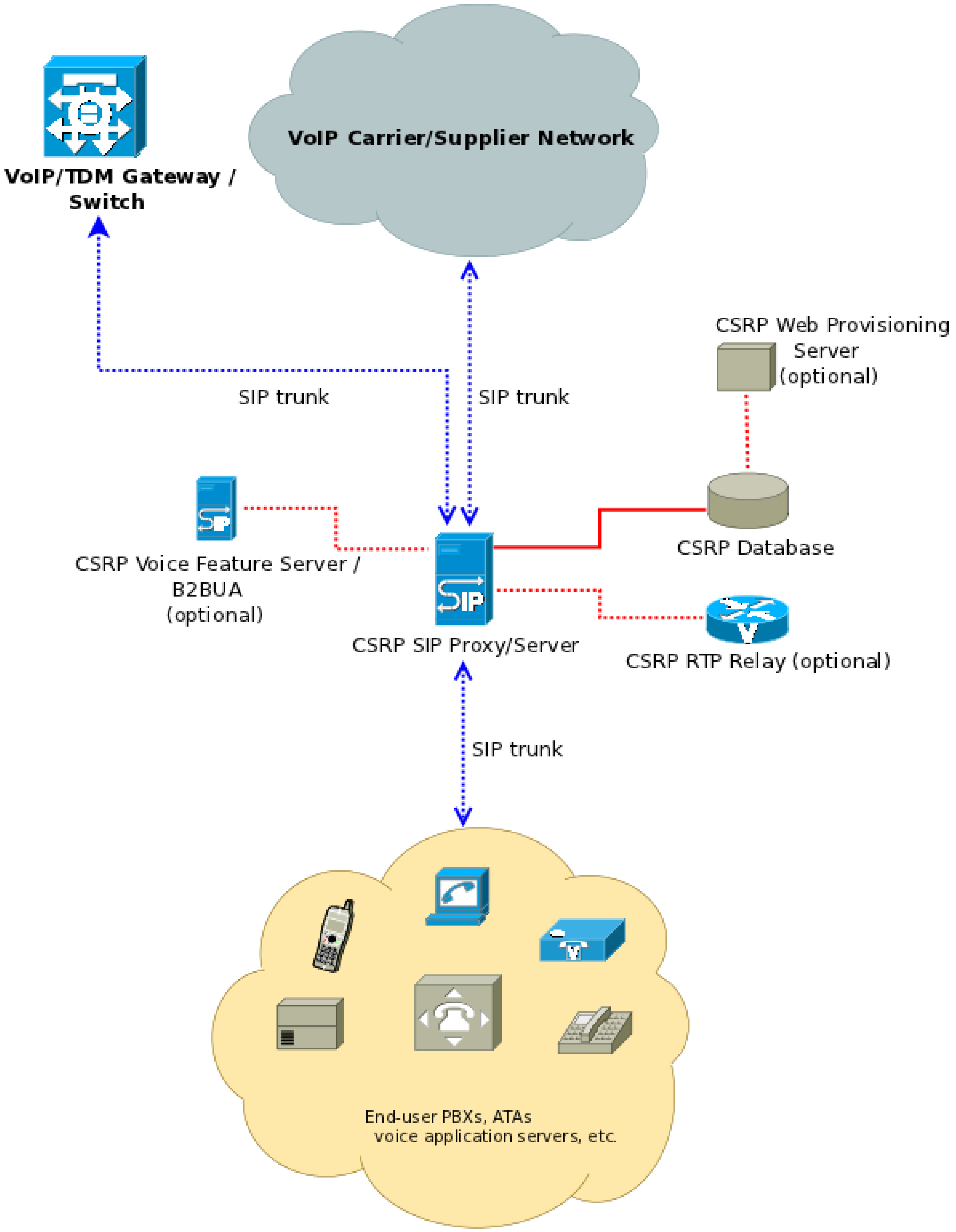
Additionally, an ITSP in the wholesale or retail origination/termination business is faced with the unpleasant choice of either paying too much for a platform whose capabilities vastly exceed their needs, or the technical burden and business cost of "reinventing the wheel", in the form of a haphazardly-constructed 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 more than adequately served. **CSRP does not aspire to be everything to everyone**. Its technical orientation, architecture and feature set is **geared for SIP trunking and wholesale termination/ origination 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 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 modifications and updates to the business layer without any need to restart the proxy.
- It maximises programmatic flexibility, allowing customisations that would not be possible in the proxy's circumscribed script execution environment.

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 backing the proxy.

The following additional components are available:

- One or more RTP media relays controlled by Kamailio, necessary for far-end NAT traversal or to maximise topology concealment.
- Web-based provisioning and administration interface.
- Integrated media server (SEMS) & back-to-back user agent (B2BUA) for optional features that require it, such as call forwarding and integrated voicemail.

1.0.5 - Hardware requirements

Hardware requirements for CSRP vary considerably with the software elements deployed (see

Section 1.0.5) and throughput expectations.

1.0.5.1 - Hardware elements

In principle, the most basic CSRP deployment (one Kamailio proxy and one database) can run on a single physical host. While this is an acceptable entry-level configuration, we strongly recommend physical separation of the proxy server and the database server.

The overwhelming majority of the workload rests on the database; the hardware resources devoted to the proxy can be comparatively light. The proxy itself, unlike most other CSRP software elements, can be virtualised.

If one or more RTP relay servers are required, those should run on dedicated physical servers as well.

1.0.5.2 - Minimal recommended hardware specifications

Recommended minimal configuration specifications:

- **SIP proxy/server**
  - Dual-core 2.0 GHz CPU (64 or 32-bit).
  - 4 GB of RAM.
  - 30 GB of system-wide disk space.
  - Gigabit Ethernet NIC.
- **Database server**
  - Quad-core 2.0 GHz CPU (64 or 32-bit).
  - 8 GB of RAM.
  - 100 GB of system-wide disk space.
  - Gigabit Ethernet NIC.
- **RTP relay server** (optional)
  - Quad-core 2.0 GHz CPU (64 or 32-bit).
  - 2 GB of RAM.
  - 10 GB of system-wide disk space.
  - Gigabit Ethernet NIC.

1.0.6 - 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:

- Additional proxies utilising the same database instance can be deployed. Failover among proxies is generally more easily accomplished on the SIP UAC side, rather than internally; IP addresses of additional proxies are set as secondary gateways in customer and vendor endpoints, or DNS SRV records can be used.

- Master-master database replication. CSRP supports full master-master replication among two **or more** database instances. Multiple proxies can be set up to bind to multiple writable database instances for maximum availability. Database instances and affiliated proxies can be located in different data centers/points of presence, allowing geographic and network diversity.

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.

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.

As well, 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.

2.0.7.2 - Pre-paid credit balance reservation

An optional fraud protection measure offered by CSRP is the "reservation" of a portion of the pre-paid credit balance upon the initiation of a call. This "reserved" credit is deducted from the customer's actual credit balance when determining whether the customer's balance has fallen below the threshold necessary to admit further calls.

The effect of this feature is to limit exposure to fraud by limiting the number of concurrent calls a customer can initiate by their pre-paid credit balance.

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. Fail-over can occur in the event that attempts to contact a vendor gateway time out, or if a vendor gateway explicitly rejects a call.

The priority metrics applied to the LCR route set can be tuned to suit the user's needs. For instance, in some jurisdictions it is preferable to return routes in descending order of prefix length (longest-prefix matching to get the most specific route), and within that route set, sub-sort the routes by ascending cost. In others, on the other hand, it is best to do the opposite. Such attributes can be flexibly defined within the CSRP LCR process.

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, as well as fallback to standard dialed number routing in the event that the LNP dip provider gateway is out of service.

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.

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 - Fail-over 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.

2.0.10 - Automatic CDR export

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

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.