Virginia Information Technologies Agency
VITA
Infrastructure Sourcing Strategy

*Report on the Comprehensive Infrastructure Agreement to the Virginia Information Technologies Agency – March 2015*

# PRICING STRUCTURE REVIEW

The Virginia Information Technologies Agency (VITA) provides IT infrastructure services to executive-branch agencies throughout the Commonwealth of Virginia, primarily through a contract with Northrop Grumman. In anticipation of contract expiration in 2019, and with recognition of the complexity of change in such a large shared services environment, VITA is currently evaluating sourcing strategies to better align with current best practices and future customer requirements. Toward that end, VITA has commissioned Integris Applied, an IT sourcing advisory firm with focus on the public sector and next-generation sourcing models, to assess the current environment and develop a long-term strategy.

This report is provided by Integris Applied to the Virginia Information Technologies Agency (VITA), its customers, and the Commonwealth of Virginia at large. It is focused on the current pricing structure within the Comprehensive Infrastructure Agreement (CIA), and how that structure may be supporting or inhibiting the Commonwealth's goals. Future reports will augment these findings with additional focus on the pricing rates themselves and contractual terms and conditions.

# Table of Contents

# 1. Executive Summary

## 1.1 Findings Overview

Technology services sourcing agreements are complex. They are designed to formalize relationships, define services and pricing, and to provide service provider and buyer alike the tools necessary to deliver services and evolve those services as market offerings or customer requirements change over time. Such agreements must allow a service provider to make a profit at a price that is of value to the buyer. They must allow the customer to increase or decrease volume easily, with understanding of how charges will adjust.

A poorly structured, vague, or opaque agreement will inevitably promote the wrong behaviors, reduce likelihood of achieving initial and evolving goals, and cause a breakdown in relationships amongst all parties – service provider, buyer, and end customers.

As the Commonwealth of Virginia (COVA) contemplates its next steps for technology and sourcing strategies, an understanding of its current pricing structure, how that pricing structure compares to the marketplace, and how that structure is helping meet both enterprise and agency goals is necessary. This paper provides insight into the most critical aspects of the Comprehensive Infrastructure Agreement's (CIA) pricing structure.

While the CIA contains many elements of market standard contracts, the current pricing structure is not optimal for the following reasons:

- Does not efficiently align risks and rewards between the service provider and buyer;
- Does not provide ideal incentives for service evolution and consolidation;
- Does not address fixed and variable cost appropriately; and
- Does not clearly define the financial responsibilities of both parties.

While any future pricing structure must be negotiated with potential service providers, certain terms now considered standard in the marketplace are missing from the 10 year old CIA. Including these terms in future pricing structures will provide COVA and its potential service providers with greater transparency, more flexibility in the transitioning of services, improved agility when looking for new services, and a more equitable sharing of the risk inherent in the delivery of services.

This report is the product of Integris Applied's review of the CIA contract documents and invoices, interviews with leadership and subject matter experts within VITA (finance, sourcing, and service delivery), interviews with customer service recipients and experience creating and negotiating contracts of similar size and scope. As such interviews and data gathering are ongoing, further inputs may augment this report over the course of this engagement.

# 2. Introduction to IT Services Pricing

## 2.1 Principles

An appropriate pricing structure is critical to the long-term success of any IT services sourcing agreement. The pricing structure drives incentives for both the customer and the service provider – impacting customer choice and budgets, and service provider investment and profitability. There are several ways to structure agreements, each of which may be successful in the right context. In any environment, it is important to keep the following properties in mind when developing a pricing structure:

1. **Predictable:** To the greatest extent possible, customers should be able to forecast charges ahead of time; changes in pricing that occur over time should not be a surprise.
2. **Manageable**: The pricing should not be so complex that it is needlessly difficult to administer. If quantities of work or equipment in the environment must be measured, then those quantities should be as easy and transparent as possible to measure.
3. **Fair:** The service pricing must be a reasonable proxy for a services provider's underlying costs and should adequately recover those costs. Additionally, to the extent possible, the party that causes any incremental cost should bear that cost.
4. **Incentives:** All pricing structures will incentivize certain behaviors and discourage others. The goals of the sourcing program must be kept in mind when considering the behaviors that might be driven by a pricing structure. For example, a goal to encourage server consolidation might include reduced cost at a centralized data center.
5. **Flexible:** As consumption moves up and down, the charges should also adjust. Technology is an evolving industry, and the ability to turn down an old service to turn up a new service is one of the benefits of an efficient IT sourcing agreement. Such adjustments may include minor volume changes month to month, significant scope additions, reductions, or terminations, and ability of large service providers to re-deploy investments.

These principles can be in conflict with each other, and therefore the right balance must be found between them, based on the goals of the sourcing relationship. If a change to the structure adds administrative overhead but creates better proxy for cost, is it worth making? What if a change drives a more desirable set of incentives but it makes the pricing less predictable? The specific details of any tradeoffs will need to be considered when building and negotiating the pricing structure.

## 2.2 Typical Approaches

There are a number of different pricing approaches that can be used in services agreements. Primary constructs include:

1. Fixed price (or deliverables-based)
2. Cost plus
3. Time & Materials
4. Unit Rate (Price * Quantity)
5. Base Charges (Base Resource Unit rates) with ARCs & RRCs

Each approach has benefits and downsides, and most large contracts have some elements of several or all approaches. By far the most common methodology in IT infrastructure services contracts is Base Charges with Additional Resource Charges (ARCs) and Reduced Resource Credits (RRCs) (abbreviated

as Base + ARC/RRC). This methodology uses metrics called Resource Units (RUs) such as Servers, Desktops, MIPS, Phone Lines, and a set of rules by which the charges vary as the volumes of the metrics move up and down. In general, the Base + ARC/RRC methodology tends to have the best mix of predictability, manageability, fairness, incentives, and flexibility of IT infrastructure sourcing.

To prevent value leakage, any approach should make clear which provisions are in the "base" (i.e., a direct flow-down from statement of work requirements, for no additional charges except volume fluctuations), which are billable projects (i.e., one-time efforts, not otherwise contemplated in base scope or RUs), and new services (i.e., ongoing work that is materially different from existing SOW requirements, and therefore requires SOW and RU additions).

## 2.3 Context for the Base + ARC/RRC Approach

The purpose of the Base + ARC/RRC methodology is to recognize the embedded fixed cost in a services environment while preserving flexibility for both parties as changes in service consumption occur, which increases leverage and flexibility for the buyer.

Traditionally, many contracts had significant (and often untouchable) fixed charges combined with single unit rates. Identifying Resource Units and bundling all charges into them increases flexibility, by pre-allocating costs – including fixed costs – to specific towers and RUs. Combined with the appropriate contract provisions, when a tower or sub-service is terminated or replaced, it should be easier to identify, remove, and reallocate the costs.

Unfortunately many contracts do not adequately address changes to towers and RUs in the event of a significant scope shift. Additionally, in a federated environment such as a state government where agencies with separate budgets are paying for the costs of shared services, a change in volume at one agency may affect the total charges in another agency. While this causes consternation, it is a mathematical outcome of what would happen in any shared environment, where distributed customers are sharing the cost of centralized infrastructure.

## 2.4 Building Base + ARC/RRC Pricing

The Base + ARC/RRC is still the recommended approach, but there are ways to more carefully address and reallocate shared costs with the appropriate contract language, service provider relationship, and central chargeback methodology.

When initially constructing a contract to use Base + ARC/RRC, the following logic should apply:

- The initial total Base Charges (i.e., the total cost of the contract at inception, based on the initial RU volumes) should represent the total cost of managing the environment assuming no changes in volumes.
- The Base Unit Rate can be calculated by dividing the total Base Charges by the initial Resource Unit volumes.
- Each Resource Unit should have an ARC/RRC rate, which represents the variable cost of adding or removing a single unit.
- The ARC/RRC rate, therefore, should be a percentage of the Base Unit Rate. The difference in the ARC/RRC rate and the Base Unit Rate represents the embedded fixed cost. Because the underlying fixed cost should not change with the single addition or removal of a unit, the ARC/RRC

rates should be symmetrical – that is, there should be no difference in an ARC or RRC rate for a single RU.

# 3. CIA Contract Pricing Structure

The existing NG agreement is a hybrid of a fixed price contract, and a Base + ARC/RRC contract. Most towers have a large fixed charge, together with a set of Resource Units that have a Baseline Resource Rate & ARC/RRC rates. This pricing structure reduces supplier risk by ensuring a greater percentage of provider fixed costs are guaranteed in the base charges. In a more market-based contract, the service provider would share more risk, providing the client more leverage when performance milestones are missed. The CIA pricing structure in turn removes some of the intended flexibility benefits of a Base + ARC/RRC methodology – no matter how much service consumption patterns change or evolve, the supplier is guaranteed what is basically an annuity greater than what is typically seen in contracts of this size and scope.

## 3.1 ARC/RRC Rates

In general, most contracts that use a Base + ARC/RRC methodology have symmetrical ARC/RRC rates, such that the charges for a specific RU are clearly delineated as fixed versus variable. The CIA contract has asymmetric ARC/RRC rates, and so an RU is highly variable when increasing, but largely fixed when decreasing. This approach is neither reflects underlying economics, nor does it support efficient consumption management and organic growth.

Setting aside the per-tower fixed management charge, some sample RUs look as follows:

| | Base Rate | ARC Rate | RRC Rate | ARC % | RRC % |
|---|---|---|---|---|---|
| Mainframe Processing - z/OS | $965.88 | $965.88 | $458.47 | 100% | 47% |
| Unix Physical Medium (3-4 CPUs) | $1,891.17 | $1,607.50 | $756.46 | 85% | 40% |
| WAN – 4.608 Mbps | $2,737.12 | $2,600.45 | $1,623.88 | 95% | 59% |
| WAN – 6.144Mbps | $3,338.11 | $3,171.43 | $1,980.41 | 95% | 59% |
| PC Support - Laptops | $21.83 | $11.58 | $6.66 | 53% | 31% |

The delta between the ARC rate and the RRC rate causes a number of issues. In the simplest case declines in volume result in somewhat smaller rate decreases, and a spread of fixed costs back to the remaining agencies.

For example, using the circuit pricing above, an agency that had two circuits initially would pay as follows:

| | Monthly Base |
|---|---|
| 2 - 4.6 Mbps circuits | $5,474.24 |
| 1 of each circuit | $6,075.23 |
| 2 - 6.1 Mbps Circuits | $6,676.22 |

If the example agency, having started with one of each circuit, were to increase bandwidth to two Mbps circuits, it would receive a 4.6 Mbps circuit RRC and a 6.1 Mbps circuit ARC, as follows:

| | |
|---|---|
| Starting Base Charges | $6,075.23 |
| 4.6 Mbps RRC | -$1,623.88 |
| 6.1 Mbps ARC | $3,171.43 |
| **Total New Charges** | **$7,622.78** |

The final charge is about 14% higher than it would be if the agency were to have started with two of the larger circuits. The problem would be exacerbated if the agency's utilization reduced, such that it would be prudent to downsize:

| | |
|---|---|
| Starting Base Charges | $6,075.23 |
| 4.6 Mbps ARC | $2,600.45 |
| 6.1 Mbps RRC | -$1,980.41 |
| **Total New Charges** | **$6,695.27** |

In attempting to move from one of each circuit to two smaller circuits, the agency's total charges have increased approximately 10%, despite consuming less service.

In the case of a move between tiered RUs, the small RRC and large ARC results in more marginal cost to the COVA, reducing risk for the service provider and creating a disincentive to assist COVA with the evolution of services (such as moving from a mainframe to a distributed server environment). CIA contains a clause (5.3.9 of Schedule 10.1) that is intended to counteract the effect of asymmetrical ARC/RRC rate.

In summary this clause states that the parties will make adjustments if ARCs and RRCs are caused by movements between tiers rather than actual reductions and increases in consumption. A clause such as this is, however, difficult to administer; it requires an accurate initial inventory, tracking of current inventory, and tracking the reason for every change going back to contract inception.

As noted above, in the case of moving between service type, or evolution of services (e.g., migrating from mainframe to server, moving to a more virtualized server from physical, or upgrading WAN circuits) this asymmetric ARC/RRC, together with the separate per-tower fixed charges not tied any service delivery metrics, creates a significant switching cost, which can undermine what would otherwise be healthy business cases for the evolution of IT services, slowing adoption of new technologies.

## 3.2 Service Definition

Any pricing structure is dependent upon clear definition of the services that will be consumed. This helps to differentiate what is included in the base charges, versus what is considered a one-time chargeable project or a new service altogether. This task requires careful coordination between finance and delivery leads when defining scope, and the definitions are never perfect, but clear definitions allow for effective evolution of services (i.e., forming a basis for calculating what charges decrease as the new charges increase). Having clear definition of services helps prevent incremental charges in the form of unexpected projects (when the parties have misaligned expectations about what is and is not included in the base).

The most common tools to create better definition of the linkage between the services and charges are:

- Services Tier Matrix (a contract attachment that describes the standards and options for various tiered RUs); and a
- Financial Responsibility Matrix (describes what labor, hardware, software, maintenance are included in which charges); and
- Descriptive language in the Resource Unit definitions to outline what functions a given charge is intended to compensate.

| Server | Asset Allocation: Capital / Lease Cost (1) | | | | | | |
|---|---|---|---|---|---|---|---|
| Service Towers: | Existing Client Owned /Leased Assets/Licenses (7) | Software Licenses / Transfer Fees (18) | Software Currency (17) | Future/Asset Refresh | Refresh Cycle (16) | Releases / Upgrades / Enhancements | Charging Mechanism |
| **Equipment:** | | | | | | | |
| **Server (13)** | | | | | | | |
| Application Servers | Client | N/A | N/A | SP - SV | 5 yrs | SP - SV | HSC |
| Utility Servers - Email | Client | N/A | N/A | SP - SV | 5 yrs | SP - SV | Base/Variable |
| Utility Servers - Consolidated | Client | N/A | N/A | SP - SV | 5 yrs | SP - SV | Base/Variable |
| Utility Servers - All Other | Client | N/A | N/A | SP - SV | 5 yrs | SP - SV | HSC |
| Appliance Servers/Network | Client | N/A | N/A | SP - SV | 5 yrs | SP - SV | HSC |
| Infrastructure Servers | Client | N/A | N/A | SP - SV | 5 yrs | SP - SV | Base/Variable |
| Storage (3) | Client | N/A | N/A | SP - SV | 5 yrs | SP - SV | Base/Variable |
| Consumables (e.g. tape | Client | N/A | N/A | SP - SV | N/A | SP - SV | Base/Variable (14) |
| Server Monitoring | Client | N/A | N/A | SP - SV | N/A | SP - SV | Base/Variable |
| Server Backups (4) | Client | N/A | N/A | SP - SV | 5 yrs | SP - SV | Base/Variable |
| Microfiche | Client | N/A | N/A | SP - SV | N/A | SP - SV | Pass-Through |
| Cables & Connectors - | SP - DC | N/A | N/A | SP - DC | N/A | SP - DC | Base/Variable |
| Cables & Connectors - Remote | Client | N/A | N/A | Client | N/A | Client | Retained |
| **Software:** | | | | | | | |
| **Application and Utility Servers** | | | | | | | |
| Antivirus Software | SP - SV | SP - SV | n / n-1 | SP - SV | N/A | SP - SV | Base/Variable |
| Application Software | Client | Client | n / n-1 | Client | N/A | Client | N/A |
| Application Utilities | SP - SV | SP - SV | n / n-1 | SP - SV | N/A | SP - SV | SSC |
| Compilers | SP - SV | SP - SV | n / n-1 | SP - SV | N/A | SP - SV | SSC |
| Database Management Software | SP - SV | SP - SV | n / n-1 | SP - SV | N/A | SP - SV | SSC |
| Development Tools | SP - SV | SP - SV | n / n-1 | SP - SV | N/A | SP - SV | SSC |
| Email | SP - SV | SP - SV | n / n-1 | SP - SV | N/A | SP - SV | Base/Variable |
| Email CAL | SP - SV | SP - SV | n / n-1 | SP - SV | N/A | SP - SV | SSC |
| Infrastructure Management | SP - SV | SP - SV | n / n-1 | SP - SV | N/A | SP - SV | Base/Variable |
| Middleware | SP - SV | SP - SV | n / n-1 | SP - SV | N/A | SP - SV | SSC |
| Network Software | SP - SV | SP - SV | n / n-1 | SP - SV | N/A | SP - SV | Base/Variable |
| Operating Software | SP - SV | SP - SV | n / n-1 | SP - SV | N/A | SP - SV | Base/Variable |
| Operating Software CAL | SP - SV | SP - SV | n / n-1 | SP - SV | N/A | SP - SV | SSC |
| Security Software | SP - SV | SP - SV | n / n-1 | SP - SV | N/A | SP - SV | Base/Variable |

This above example of an FRM describes financial responsibility for various components of the services, refresh and currency requirements, and the charging mechanism to recover the costs. For a sample Services Tier Matrix, see section 3.2.

In addition to the STM and FRM, a pricing structure should contain definitions of services and offerings tangential to the RUs. These topics include:

- Asset installation and configuration (especially servers)

- Asset de-installation and disposal (especially servers)
- Treatment of standard vs. non-standard options (Is the price an average of the standard and non-standard mix in the environment, or is it just the cost for everyone to receive the standard option and anything non-standard is incremental)

RU definitions in the CIA are imprecise. For example the messaging RU is defined as follows (From attachment 10.1.3A):

> *Messaging: Each messaging individual or group mailbox (excluding distribution and contact lists).*

A more industry standard RU definition might read:

> *"Email Accounts" shall be a Resource Unit. One (1) Resource Unit shall equal one discrete email identification (e.g. excludes aliases, resource accounts, distribution lists) approved by Customer and available for use. Email Accounts must be inclusive of all labor, Software, Equipment, maintenance, and systems management to support services and projects (including installation and discontinuance) required to manage, maintain, and enhance the email environments. Shared storage and backup for email is not included in the email service, but are part of Server Storage Services. Usage will be provided by running a maintained script against the directory services and providing valid counts on the 15th day of the month*

CIA does not contain an FRM or STM. While it does have some level of definition in Schedule 10.1.13 (Supplier assumptions) and 10.1.14 (Reallocation and Upgrade Matrix), the lack of a STM or FRM reduces transparency, fosters mistrust between parties, and creates significant difficulties in the evolution of the services. The industry has moved away from having a contract the size and scope of the CIA that does not include these items.

## 3.3 Asset Ownership

One of the key strategic decisions buyers must make is whether or not to include assets in the service provider pricing – that is, to transfer financial responsibility for the assets to the service provider. (Note that for purposes of this commentary, "financial responsibility" means that the provider has bundled the asset cost into its pricing, either in a fully managed service or via some other payment stream such as a hardware services charge; it does not refer to situations where a client may directly buy or lease equipment from the service provider.)

There are several reasons why clients choose to transfer financial responsibility for assets to the service provider:

- Transferring risk of technology decisions to service provider
- Simplified budgeting / capital smoothing
- Standardizing infrastructure platform
- Guaranteeing refresh

Many clients, particularly in state government, have chosen to transfer assets to the service provider. This is done because it is challenging to maintain technology standards and reliability in an environment with long-term budgeting horizons, and difficult to receive consistent capital for upgrades and refresh. Purchasing assets "as a service" allows an organization to include refresh cycles in the services model,

keeping the environment current. Further, including equipment in the deal allows a state to transfer more risk of service delivery to the provider, more fully leveraging the marketplace's technical expertise.

However, there are significant downsides with including assets in scope. Doing so limits the number of service providers that can bid (i.e., only the largest of service providers have the available capital), it makes the contract terms longer, the termination costs higher (to allow the provider to recover its investments), and at termination or expiration the assets need to be valued, bought out, and transferred internally or to the new provider. All of these items increase switching costs, and generally reduce client leverage to negotiate price and service delivery.

In addition to the primary issue of reduced leverage, including assets in scope also requires tax to be collected and billed back to the COVA; depending on how it is embedded, asset inclusion can result in a relatively fixed hardware "standard" that does not evolve quickly enough as the hardware market changes, and it removes COVA's ability to defer refresh to produce savings in a bad budget year (although of course great care must be exercised, because if this option is exercised too frequently, large total costs increases can result from an increasingly difficult to support aging environment).

The difficulty with moving back to an asset-retained position is the necessity of a one-time infusion of capital; however, in many cases making this transition can result in operating savings. It can also be the case that owning the assets can requirement more administrative oversight than having a fully bundled service.

This issue is a complex one driven by the business and capital needs of the buyer, and while there are good reasons for either including assets or excluding assets, great care must be taken to establish the right contract terms for flexibility and cost recovery. For further reference information, we have published a white paper describing decision criteria and tradeoffs for asset ownership inclusion.

### 3.4 Cost of Living Adjustment (COLA) & Price Performance

The CIA has a COLA clause that adjusts the pricing on a yearly basis as a function of a set of price inflation indices. This type of clause is common in the market, but what is unusual is that the CIA has flat pricing year over year in the before COLA is added. In general, most contracts of this type show year over year price performance (the effects of reduced hardware costs and increased efficiency), which nets an annual cost reduction.

In any static IT services environment, most service provider costs go down year over year. Automation increases efficiencies, hardware costs decrease, and environments are consolidated and standardized. Although some specific underlying cost components may increase (e.g., personnel salaries and benefits, software licensing) the net result of operational efficiencies and equipment evolution is an overall reduced cost. This is not the case with the CIA.

### 3.5 Consolidation

CIA pricing structure does not provide incentives for either Agencies or the supplier to move to the consolidated data center or standardize, and limited incentives to virtualize. The economics of administrating standardized, consolidated virtual servers are far better than those of distributed, non-standard, physical servers, excluding any one time migration costs. Approximately two thirds of in scope servers remain outside the supplier's state of the art data center. Creating a set of RUs that reflect the

economics that consolidation creates would provide incentive for further consolidation, standardization and cost savings.

### 3.6 Resource Units

In a Base Charges + ARC/RRC methodology, resource units must allow the service provider to recover its costs and earn a fair return on capital. Designing resource units to allow service provider profitability while giving the customer transparency and a fair price is an important part of a pricing structure.

There are six general rules for the design of an efficient resource unit:

- They should be a reasonable proxy for service provider variable costs
- They should be easy to measure
- They should be predictable based on customer business volumes
- They should effectively align risk and incentives
- They should minimize complexity in the environment
- They should clarify service delivery responsibility

Resource units that meet these criteria will allow the service provider to recover its costs and earn a fair return on capital be as easy as possible to operationalize, drive each party toward its goals, and not cause uncontrollable shifts in customer costs.

Because one of the main financial goals of any outsourcing agreement is to provide pricing certainty, it is important that the contract structure is setup such that for a consistent scope of services, changes in volumes are compensated only through ARCs and RRCs of the Resource Units whose volume is changing. Because these ARC/RRC charges will be the only changes in cost due to volume variations, it is important that they are well aligned with the supplier's marginal cost.

## 4. CIA Contract Resource Units

### 4.1 Storage

The CIA contract storage resource units are arranged into tiers based on the underlying hardware and service needed to support the hardware. There are two broad categories, one for SAN storage and the other for direct attached storage. Volume are measure per-GB allocated for SAN and per-GB installed for direct attached [not sure how deep to go on this topic, it's important, but also pretty technical, in general per-GB allocated is better because it aligns with business consumption rather than what the technical folks decide to install (this is really only true if we police what they allocate, although it is the case that if we find an error later it's a lot easier to de-allocate than de-install)]. The resource includes all of the necessary hardware, software, and labor to operate the storage environment. Additionally, the cost for the backup environment is included in the storage rate.

What's most common in the industry is to have SAN billed per Allocated GB, and it's normal to have all the necessary hardware, software, and labor included in the RU. By way of comparison, direct attached storage (i.e., the storage in legacy servers) is generally not separately chargeable, because it's extremely difficult to inventory, and the industry is moving more and more towards have critical applications and data resident on SAN.

Rather than having backup included in the storage charges, it's more common to separate the cost for backup into its own set of charges. The largest difference between the current CIA contract and the market position is the inclusion of backup charges in the storage rates.

One benefit of including backup costs in the storage rates is that the backup resource units (when they are separated out) are difficult to measure. They tend to be a mix of GB stored, GB written, and tapes, depend on a variety of underlying architectures that are often very difficult to report out of and aggregate together.

On the other hand, having something that can vary as much as backup requirements (as a function of backup frequency, retention policies, etc.) included in the RU can cause either underperformance in service provisioning with the service provider pushing everyone to a minimum standard offering, or open the door to frequent change orders and project charges for the implementation of the requirements of those agencies which require more backup.

In future procurements, having as large an item as backup included in the RU and not separately chargeable may be a difficulty to bidding providers who don't know the environment and may have some significant risk around both what the backup environment is and how frequently the requirements change. It's also the case that things that are split out into RUs tend to develop relatively accurate inventories, while those that are included in an RU often have their inventories degrade overtime. Whether a decision is made to split out the backup charges or not it may be beneficial to launch an enterprise inventory activity of the backup environment in advance of any future procurements (to support accurate provider solutioning and pricing).

## 4.2 Server

Server charges under the NG contract are separated in the following manner:

- Unix and Intel
- Physical or virtual
- Tiers based on server size (number of processors).

All of the hardware, software, and labor for support of the servers are included in the RU. DR for servers is in a separate RU. Server standup and shutdown are compensated via a project charge. There is a concept of a non-billable server type that NG uses in its delivery of services (DNS, DHCP, the servers that support backup, etc.)

The market standard approach is to split the charges for hardware and support into separate RUs. In general the level of support for a single instance is not dependent upon the size of the hardware it is running on. Hardware size tends to depend heavily on the applications being run, and hardware itself tends to evolve quickly compared to the services contract cycle. Additionally in the Intel space at least, hardware has been commoditized.

In the services space, the support needs for a given instance vary depending on the instance's purpose. A production instance in a life and limb application requires faster incident response and better uptime than a dev/test instance used in a non life and limb application.

These factors together lead to an industry standard structure where the support charge per instance is divided into Intel/Unix, then further divided based on the level of support that that instances receives. These levels of support are documented in a Services Tier Matrix that ties together service delivery responsibilities, SLAs, and pricing.

S = Standard: service provided within tier
O = Optional: can be added at additional cost as described in **Attachment 4-A**
X = Required element for entry to tier
- = Not Available

| Offerings: | Platinum | Gold | Silver | Bronze |
|---|---|---|---|---|
| **Service Levels** | | | | |
| Availability | 99.95%/99.9% | 99.9%/99.8% | 99.85%/99.75% | 99.75%/99.65% |
| Incident Resolution Time - Severity 1 | 1h | 2h | 4h | 6h |
| Incident Resolution Time - Severity 2 | 2h | 3h | 6h | 16h |
| Root Cause Analysis Delivery | 10d | 10d | 10d | 10d |
| Successful Recoveries | 6h | 24h | 48h | 72h |
| | | | | |
| **Monitoring** | | | | |
| Network Monitoring | S | S | S | S |
| Hardware | S | S | S | S |
| Operating Level (OS) | S | S | S | S |
| Database | S | S | S | O |
| Application Software | S | S | O | O |
| Middleware Processes | S | S | S | O |
| Capacity Utilization Reporting | S | S | * | * |
| Capacity Management | S | S | * | * |

For hardware, the charges are generally not an amount defined up front, but instead based on the underlying cost of hardware and hardware maintenance. This hardware services charge allows flexibility in hardware choice, and the use of competition to keep hardware prices down. In situations where it is necessary to drive standardization, it's still not necessary to "bake in" the hardware, a central organization can chose the standard models, which can be updated as often as necessary, and lists them on the services catalog.

Server software generally has two treatments. Operating system, antivirus, monitoring tools, etc. are generally embedded in the server rates (and defined in a software exhibit which ties back to the financial responsibility matrix.) Software which is more closely aligned with agency applications (databases and middleware) and the applications themselves are generally retained, or priced in a software services charges (that functions largely as a pass-through)

Currently database support is included in the server support charge, but separating it out into its own charge, minimizes cross subsidization between agencies that consume more and less database services, and further allows the price to adjust as more or less database support is consumed.

The pricing structure in CIA does not provide the level of transparency and predictability seen in most agreements of this size and complexity. Accordingly, the parties are denied a tool that would allow for improved decision-making on cost and the allocation of resources.

## 4.3 Mainframe

The mainframe tower includes the support of two separate environments, the IBM mainframe and the Unisys mainframe. It contains a variety of charges, for storage, tape, etc. but the majority of costs for mainframe service delivery are contained in the mainframe computing charges. These mainframe compute charges are assessed for blocks of installed MIPS, 425 MIPS for the IBM side, and 50 MIPS for the Unisys side.

The industry standard for mainframe charges is CPU hours, although this methodology is better developed and more common with IBM platforms that Unisys platforms. CPU hours uses the IBM mainframe's time accounting system to sum the actual time used on the processors by the various jobs

running on the mainframe, excluding system overhead (provider run jobs for security, scheduling, auditing, etc.), and normalized to a consistent platform. CPU hours is a lot more flexible charging mechanisms than MIPS, with cost being added and removed smoothly as volume is increased and decreased.

MIPS tends to be a better proxy for supplier cost, and it also tends to be easier to administer, at the cost of being far more fixed than CPU hour (costs come out as a large step function, when they come out at all). It also puts consumption management functions back on the client (regardless of what the SOW says), the provider cannot make decisions about how much headroom to maintain, because each additional CPU is a billable item, at best they can provide reporting for client consumption management.

A move to the industry stand CPU hours would provide increased pricing flexibility and improved incentives at the cost of increased administrative overhead.

## 4.4 EUC

End User Computing charges in the contract consists of two parts, a support RU per Desktop, Laptop, or Tablet, and a hardware RU. The OS licensing is included in a separate Microsoft charge. This structure is comparable with the market, save that like Server, when the hardware charge is split out, it is generally a charge based on the underlying cost of the chosen asset to give client control of the hardware choice and frequency of standards update.

## 4.5 Messaging

The messaging tower has one primary resource unit: the mailbox. This is normal in the industry. The count of mailboxes excludes distribution lists, but includes group mailboxes (which are sometimes excluded). The mailbox RU generally includes the necessary labor, software, hardware to run email, save that the Exchange licensing is in a separate Microsoft charge. The CIA also offers upcharges for additional services, such as additional security or archiving.

## 4.6 Service Desk

The Service Desk tower's primary RU is per supported end user computing device. The ideal approach (and industry standard for large sourced environments) is to use authorized users or supported devices. Service providers sometimes express a preference for a "per call" RU because they size their staffing based on numbers of calls and time to resolve. However, when service desk and incident management are the service provider's scope, using a per call metric does not incentivize the provider to reduce user contact to the service desk via improved services and "shift-left" self-resolution tools.

## 4.7 Cross Functional Services

The CIA cross-functional services tower does not have any resource units, but rather a set of fixed charges. Having separately-priced cross-functional services has not been common in the market historically (they are generally embedded in the other tower charges), and it is only recently becoming common. The industry has not standardized on a methodology for charging these services yet, but the three most common are FTEs (billable cross-functional time multiplied by the appropriate rates), transactions (e.g., incident tickets, assets managed in CMDB, total changes processed through CAB, etc.), and a set of proxies for the environment managed (e.g., cross functional services provided for X

servers, Y desktops, and Z circuits). All of these methods have benefits and downsides, but all three provider better flexibility, incentives, and economic alignment than a single, large fixed charge.

## 4.8 Data Network

The data network services in the environment are provided across several contracts. NG is the required provider for executive branch agencies, but a Verizon contract is also in place, which out-of-scope agencies can used to buy unmanaged services.

### 4.8.1. Local Area Network (LAN)

Wired LAN services charges are primarily covered by the LAN Port RU, which is an industry-standard approach. Active LAN ports are an ideal metric because it encourages service providers to build the infrastructure in an efficient way. However, it is often difficult to count active LAN ports, particularly in a first generation sourcing environment where inventories may need to be completed, service request processes established, and with legacy (unmanageable) equipment still in place. The NG contract addresses the counting issue by summing other RU volumes in the environment connected to the network (e.g., servers, PCs, etc.). While this approach has the advantage of simplicity, it also has the disadvantage of not addressing all of the ports in the environment (such as conference rooms), discourages a move to wireless, and make it difficult to build out a LAN in a new site or agency where little or no in-scope PCs may exist. In some environments, service providers may have significant unrecovered LAN costs, which can lead to undesirable outcomes such as costs allocated to other RUs and requests to negotiate a new RU or lump-sum payment to address the uncountable ports.

In fact, in the current environment the service provider is charging customers separately when they are not able to recover LAN costs via PC/printer/server counts. For example, if an agency were to install a number of new network-connected cameras or conference room phones, requiring a new switch, the service provider requires the customer to buy the switch outright. Yet, the service provider is able to count ports that are used over the course of the month for purposes of service level reporting, because all sites have been refreshed with modern equipment. This same active monitoring approach could be used for counting the billing unit. There are some sites that cannot be monitored (since they are not the MPLS network), but those are small sites and active ports could be counted as switch capacity.

In a future model, it may be possible to use an updated inventory and the service request process to maintain the active port inventory. When defining an active LAN port RU in a contract, a key part of the definition should be “active and authorized for use by the customer.” The word “active” means that it is live and working, and “authorized” means that the customer wants it to be live and working. If either of those conditions is false, then it should not be charged. In other words, if a starting inventory of active ports exists, then the service request process can be used to manage the count: a user request to activate a port initiates a charge, and a user request to de-activate removes a charge.

Either of the recommended approaches to count active LAN ports (rather than assets that might connect) would be more effective for all parties.

Wireless access points are charged per device, which is also a typical metric. In many environments, and as seems to be the case here, the choice of wired and wireless RU can make it difficult to justify a move to wireless in existing or new environments.

### 4.8.2. Wide Area Network (WAN)

WAN has costs for circuits, routers, maintenance, and support labor combined into single per circuit and per managed router RUs. A more typical approach is to separate the circuit charges from all the other charges. This is because circuit charges should be 100% variable; in other words, when a customer requests a circuit to be disconnected (or replaced), then the full cost of the old circuit should disappear from the bill. This separation of costs also allows for easier cost comparison (and in some cases procurement) of what is largely a commodity service. Other costs, such as router maintenance and the enterprise support processes are less variable, and can be bundled into a per site or per router charge. The current structure represents that there may be ongoing fixed costs associated with a circuit, even when sites are completely disconnected. Further, the asymmetric ARC/RRC rates make circuit upgrades more expensive and difficult to justify.

Prior to Mod 60, all WAN and Internet backbone costs were included in LAN Port charges. Following Mod 60, circuit and router charges were separated from the LAN, and upgrades to the enterprise internet connectivity were charged separately.

There are inefficiencies in the MPLS RUs, including a lack of enough bandwidth choices, inefficient pricing jumps between capacities, lack of options for quality of service (QOS) without pricing changes, and costs of primary offerings causing in-scope agencies to use less robust connections.

## 4.9 Voice Network

Voice services in the agreement consist almost entirely of VOIP-related RUs. However, approximately 2/3 of the originally planned scope was transitioned to NG VOIP (UCaaS) services, and even that did not represent the entire in-scope footprint.

The remaining legacy sites are using a Verizon contract to buy traditional voice services and receiving maintenance on PBXs and key systems from NG. This pricing approach discourages continued transition to VoIP services, because each new customer would pay for seats, and the enterprise cost for the traditional voice does not reduce automatically when customers upgrade to VOIP.

Long distance charges are included within the VoIP rates, but capped at an amount that has not been exceeded. It is industry-standard to include long distance in hosted VoIP offerings, but buyers should avoid caps.

There are a significant number of VOIP RUs, covering the phone sets and various feature options. That does allow customer choice, but it may be possible to simplify the number of RUs by bundling features.

Traditionally, one of the difficulties in migrating from legacy voice environments to VOIP was the difficulty in justifying the cost. The issue is not just the potentially significant cost of an infrastructure investment, but also the complexity in calculating a business case to address reduced voice circuits, long distance minutes, new data network capacity, etc. One of the ways clients have simplified that is by transferring the risk (and benefit) of transformation to the service provider, by having a simple per port charge for all of voice services, whether legacy or VOIP. The charge to the customer may be different for different types of port or handset, but those price differences are based on feature differences rather than underlying. In other words, the customer cares that she can get a dial tone with some set of features (perhaps the ability to place calls on hold or transfer), and does not care whether the call was delivered on the PSTN or an internal IP network.

# 5. Conclusion

The CIA does not contain an optimal pricing structure for the Commonwealth. While it includes many industry standard terms on its face, a further examination reveals that the CIA lowers service provider risk at COVA's expense and provides disincentives for standardization, further consolidation and service evolution. While no pricing structure can, alone, account for all components of service delivery, a good pricing structure reduces the risk of dispute and provides a framework for constructive dialogue between supplier and provider on pricing and terms. Pricing structure is also one piece of broader services agreements. Any pricing structure must be viewed through the lens of a Master Services Agreement (MSA), statements of work, and the financial terms and conditions of the MSA. These topics will be discussed in future deliverables.