System Architecture for Server-Based Network-RTK using Multiple GNSS

Samsung LIM & Chris RIZOS, Australia

Key words: GNSS, NTRIP, RTCM, Server-based RTK, Network-RTK

SUMMARY

Improved accuracy and reliability of the real-time user coordinates is expected if GNSS data is processed using the network-RTK technique. Some implementations of conventional network-RTK, e.g. Virtual Reference Station (VRS) techniques, require bi-directional communications via GSM/GPRS since the network server transmits network-corrections to a user, while the user receiver is required to transmit its (approximate) coordinate information to the server. The data transmission from the user is kept to a minimum because the server needs the user's coordinate information only for "initialization" or "handover". On the other hand, a server-based network-RTK requires raw measurements from the rover/user receiver in order to compute the rover's coordinates. There are several commercially available network-RTK systems, however server-based techniques are relatively new in GNSS surveying. In GNSS geodesy, server-based network processing has always been the norm, though rarely in real-time. With the move of the IGS towards a "real-time future", the distinctions between server-based network-RTK implementations for survey users and for geodesy will be blurred.

This paper describes a server-based network-RTK design that uses distributed servers to compute a rover's coordinates in the required reference system by taking advantage of existing GNSS reference network infrastructure, instead of broadcasting corrections or data to users and placing the onus of obtaining a final solution on clients and their equipment. The system architecture of the proposed server-based network-RTK is described in this paper. The main consideration is that distributed-computing is necessary to cope with simultaneous requests from hundreds of clients. Distributed-computing allows computers to efficiently communicate amongst themselves and individually process data.

System Architecture for Server-Based Network-RTK using Multiple GNSS

Samsung LIM & Chris RIZOS, Australia

1. INTRODUCTION

Continuously Operating Reference Stations (CORS) of Global Navigation Satellite Systems (GNSS) are valuable infrastructure for high accuracy applications in surveying, mapping, navigation and geodesy. CORS networks are being established at an ever increasing rate around the world, and this infrastructure will be used for long-term geo-scientific studies as well as provide the basis for Real-Time Kinematic (RTK) positioning and augmentation services (including Assisted-GNSS or A-GNSS).

The standard mode of commercial GPS carrier phase-based positioning for surveying and mapping applications is the baseline approach where one GPS receiver is located at a point of known coordinates and another (the 'rover') receiver would have its coordinates determined relative to the fixed reference station. In this mode the longest practical baseline length is typically between 10-20km, and varies with ionospheric activity, satellite geometry, and the effects of other residual biases.

In the 'network mode' the CORS inter-receiver distances can be relaxed to 50-70km, making it an economical technique that takes advantage of CORS coverage across large areas. In early-2003 the Department of Lands, in the state of New South Wales (NSW), Australia, initiated a project to "…establish a multi-reference differential GPS network capable of providing suitable equipped users with centimetre level position accuracy in real time across the greater Sydney metropolitan area." This network, known as "SydNet" (see Figure 1), was also intended to be a testbed for University of New South Wales (UNSW) research into CORS network-based algorithms, technologies and concepts.

At the time of writing of this paper, SydNet consists of seven stations in the Sydney basin area plus four more regional stations at Bathurst, Newcastle, Nowra and Gouldburn, with plans for two more at Port Kembla and Wyong in the 2008-09 financial year. Over the next few years the network will extend across the state of NSW. In addition, UNSW hosts several CORS receivers, including a GIOVE-A tracking receiver.

The extension of SydNet, to eventually become "NSWNet", is partly driven by the investment in GNSS CORS infrastructure which is being made possible by the investment from the Federal Government's National Collaborative Research Infrastructure Strategy (NCRIS) for geo-scientific research known as "AuScope". Over 100 new GNSS CORS will be established across Australia over the next three years.

Current CORS real-time augmentation systems can be described as being "generation 1", as the RTK services that are provided to a small number of subscribers on a "use-as-is" basis, without guarantee of service. CORS network hardware and software systems are going to face a number of challenges in the coming years, including the need to track new GNSS signals, generate new RTK correction messages, integrate free real-time data streams from the International GNSS Service (IGS) and other CORS networks, develop improved algorithms to support ultra-high accuracy geodetic applications, and develop new CORS-based services.

All of these challenges will require a paradigm shift away from service providers being the government agencies that invest in the CORS infrastructure (receivers, communication links, servers, etc.) to a variety of new architectures and business models that can take advantage of the opportunities offered by improved CORS network infrastructure.

Research challenges range from designing the appropriate IT components and GNSS CORS receiver technology, and the new baseline and network algorithms utilising the new GNSS signals, new data processing models based on different configurations of CORS and user GNSS receivers, leading to new RTK services, delivered by a variety of new wireless communication links, incorporating new value-added services such as server-based processing, which more closely link RTK to real-time geodesy operations and fundamental geo-scientific applications.

Figure 1: SydNet CORS network in New South Wales, Australia.

This paper describes a server-based network-RTK design that uses distributed servers to compute a rover’s coordinates in the required reference system by taking advantage of existing GNSS reference network infrastructure, instead of broadcasting corrections or data to users and placing the onus of obtaining a final solution on users. However, the proposed system requires two-way communications between servers and clients which makes it possible to provide conventional RTK services as well as post-processing services (see Figure 3).

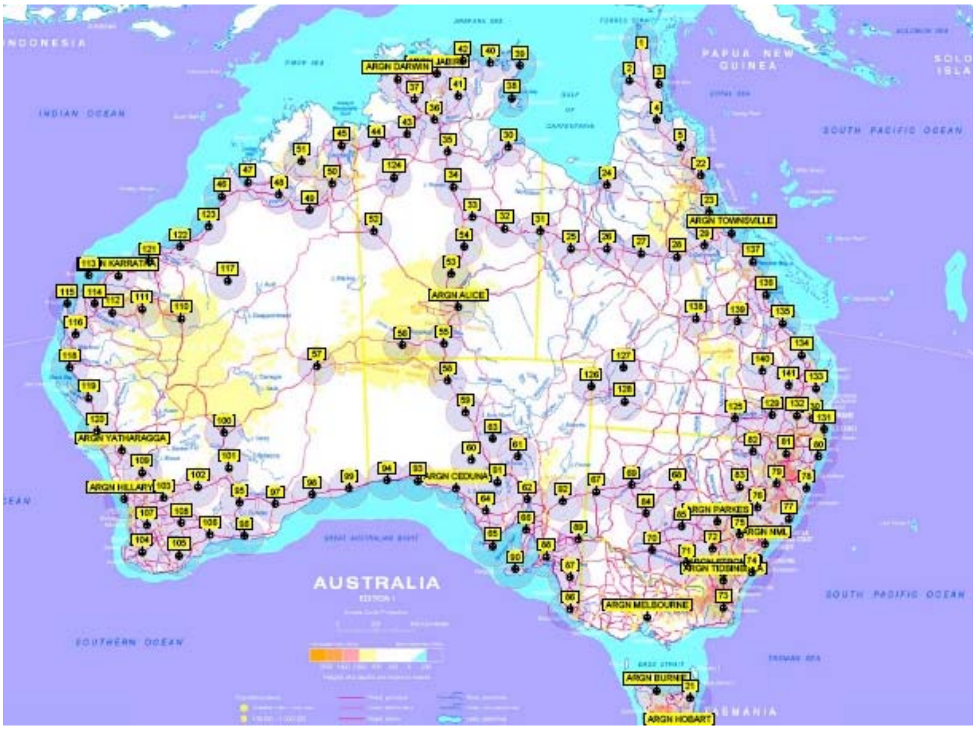

Figure 2: Planned AuScope GNSS CORS receiver sites.

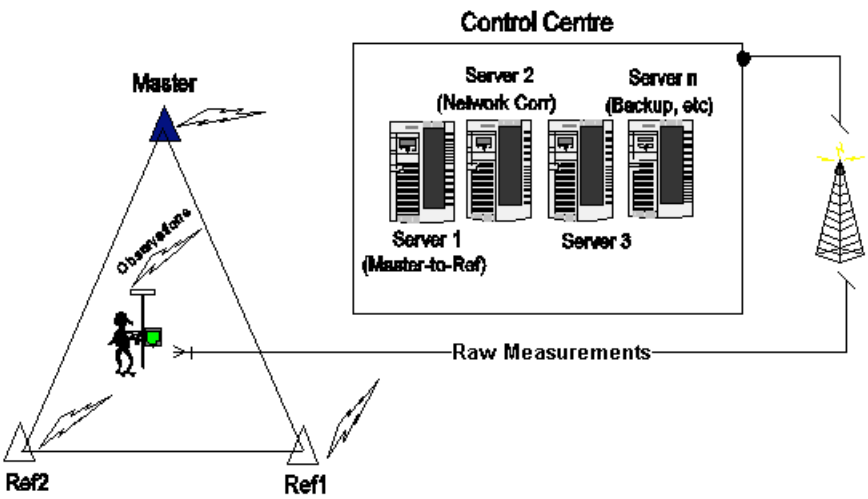

Figure 3: Server-based Network-RTK.

The main consideration is that distributed-computing is necessary to cope with simultaneous requests from thousands of clients. Distributed-computing allows computers to efficiently communicate amongst themselves and individually process data, which is different from networked-computing. Therefore the following items are to be considered at the development phase of the proposed system.

- Huge volumes of data are involved (e.g. in NSW, 100 CORS, each tracking 30 satellites at 1Hz and 1,000 or more simultaneous users). Twelve month storage into the database is recommended for post-processing users.

- Quality control and GNSS data integrity are critical factors in the take-up of this positioning technology for high precision applications (e.g. machine guidance).

- Most commercial CORS management systems store GNSS data in file-based structures, whereas the DBMS structure provides more versatile and secure storage options.

2. SERVER-BASED NETWORK-RTK

The concept and technique of network-based (differential) GPS carrier phase positioning was first introduced by Wanninger (1995). The technique utilises at least three reference stations from a local CORS network as shown in Figure 4. One of the reference stations can be treated as a 'master' station, usually the nearest to the roving user receiver.

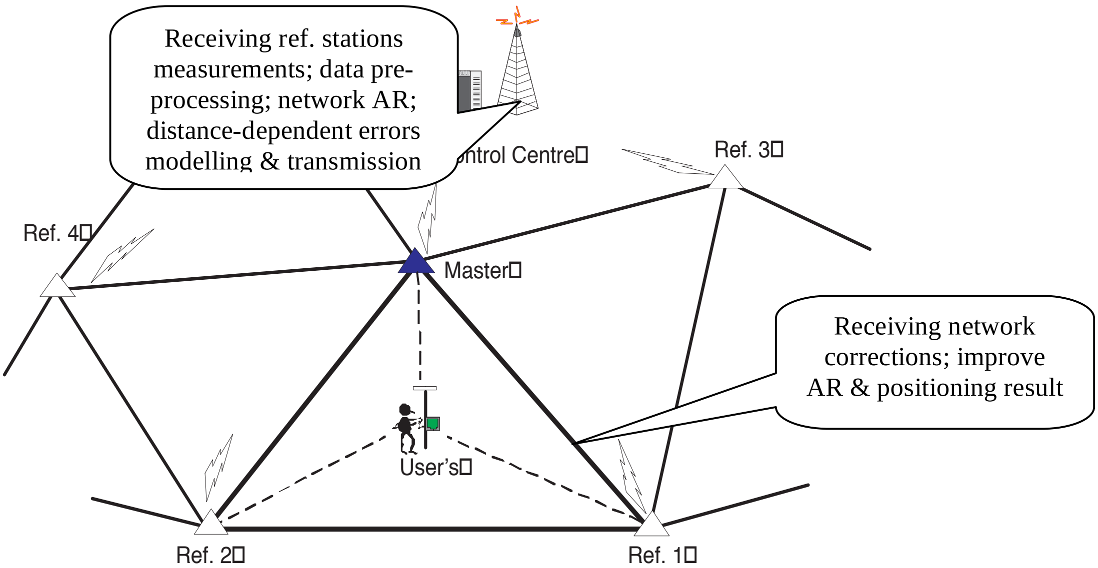

Figure 4: Network-RTK concept

The reference stations, using observations to the individual satellites, are able to estimate the atmospheric delay and orbital error (i.e. the source of distance-dependent errors) within the network on a satellite-by-satellite basis. The measurements from each reference station are sent to a control centre where these measurements are combined into a model. The control centre is responsible for the basic data pre-processing, such as cycle slip detection and repair, applying an a priori tropospheric model, applying the antenna calibration model, and resolving the network ambiguities (from 'master' to the other surrounding reference stations). The following assumptions have been made in implementing the network-based technique (discussed in the context of double-differenced measurements):

Assumption 1: Once network carrier phase ambiguities are resolved, the *residuals* contain the (remaining) *correlated* and *uncorrelated* errors within the network.

Assumption 2: *Correlated errors* are ionospheric delay, tropospheric delay and orbital error. They have spatial and temporal characteristics. The errors can be spatially modelled using the network-based approach.

Assumption 3: *Uncorrelated errors* are multipath effect, antenna offset (and variations) and measurement noise. These are station-dependent and thus cannot be mitigated by the network-based approach. The effects are minimised by calibration (Wanninger & May, 2001; Park et al., 2004), careful site and hardware selection, and the application of special techniques (Wübenna, et al., 1996).

Assumption 4: Correlated errors can be partitioned into *dispersive* and *non-dispersive* components. *Dispersive error* is related to the ionospheric delay, and is frequency-dependent. *Non-dispersive error* is related to tropospheric delay and orbital error, which is frequency-independent.

2.1 Residual Vectors for Reference Stations

Assume a CORS network of n reference stations $R_1, R_2, \ldots, R_n$; unknown variables in a set of double-differenced measurement equations are ionospheric delay I_1 (from the primary frequency f_1), tropospheric delay T, and two integer ambiguities N_1 and N_2 (from the two frequencies f_1 and f_2 respectively). Multipath effect is ignored here as it is assumed to be low or negligible at CORS that have clear line-of-sights to all satellites. The equations can be reformulated with unknowns I_1, T, and the wide-lane ambiguity $N_{1,-1} = N_1 - N_2$. The wide-lane ambiguity can be easily determined as it has a relatively long wavelength (~86.2cm in the case of GPS), which is important input information for resolving N_1 and N_2.

The primary ambiguity N_1 (and then N_2) can be fixed by rounding the combination of $N_{1,-1}$ with the ionosphere-free ambiguity $N_{77,-60} = 77N_1 - 60N_2$ (Goad, 1992), or by the least-squares ambiguity decorrelation adjustment (LAMBDA) (Teunissen, 1995), or by a Kalman filter technique described below. The phase range equation for $N_{77,-60}$ can be rewritten in order to separate the ambiguity into $N_{1,-1}$ and N_1:

$$\lambda_{77,-60}\phi_{77,-60} = \rho + 60\lambda_{77,-60}N_{1,-1} + 17\lambda_{77,-60}N_1 + T + \varepsilon \tag{1}$$

where ρ and ε represent true range and the unknown error propagation, respectively, $\lambda_{77,-60}$ is the wavelength of the ionosphere-free frequency combination $f_{77,-60} = 77f_1 - 60f_2$, that is, $1/\lambda_{77,-60} = 77/\lambda_1 - 60/\lambda_2$, and $\phi_{77,-60}$ is the ionosphere-free phase range $\phi_{77,-60} = 77\phi_1 - 60\phi_2$. The residual tropospheric delay can be approximately expressed as (Zhang & Lachapelle, 2001):

$$T = T_z / \sin z \tag{2}$$

where T_z is the relative tropospheric zenith delay (RTZD) and z is the elevation angle. RTZD and N_1 will be estimated in real-time with a Kalman filter based on the model (Hu et al., 2003):

$$\begin{aligned} y_k &= H_k x_k + u_k \qquad u_k \sim N(0, R_k) \\ x_k &= \Phi_{k,k-1} x_{k-1} + v_k \qquad v_k \sim N(0, Q_k) \end{aligned} \tag{3}$$

where the observation vector y_k is given by:

$$y_k = \begin{bmatrix} \lambda_{77,-60}\phi_{77,-60}^{1,2} - \rho^{1,2} \\ \lambda_{77,-60}\phi_{77,-60}^{1,3} - \rho^{1,3} \\ \vdots \\ \lambda_{77,-60}\phi_{77,-60}^{1,n} - \rho^{1,n} \end{bmatrix} \tag{4}$$

with superscripts represent satellite pairs for differencing, and the state vector x_k is:

$$x_k = \begin{bmatrix} T_z \\ 17\lambda_{77,-60}N_1^{1,2} + 60\lambda_{77,-60}N_{1,-1}^{1,2} \\ \vdots \\ 17\lambda_{77,-60}N_1^{1,n} + 60\lambda_{77,-60}N_{1,-1}^{1,n} \end{bmatrix} \tag{5}$$

and the design matrix H_k is:

$$H_k = \begin{bmatrix} 1/\sin(z)^{1,2} & 1 & 0 & \cdots & 0 \\ 1/\sin(z)^{1,3} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1/\sin(z)^{1,n} & 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{6}$$

The state transition matrix and corresponding covariance matrix are defined as:

$$\Phi_{k,k-1} = \begin{bmatrix} e^{-\Delta t/\tau} & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \quad Q_k = \begin{bmatrix} \frac{\tau}{2}(1 - e^{\Delta t/\tau})q^2 & 0 & \cdots & 0 \\ 0 & 1e^{-16} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1e^{-16} \end{bmatrix} \tag{7}$$

where q is the variance of RTZD noise for the correlation time τ; and Δt is the sampling rate. A residual-based adaptive procedure is proposed to improve the reliability of the Kalman filter:

$$
\begin{aligned}
x_{k,k-1} &= \Phi_{k,k-1} x_{k-1,k-1} \\
P_{k,k-1} &= \Phi_{k,k-1} P_{k-1} \Phi_{k,k-1}^T + Q_{k-1} \\
J_k &= P_{k,k-1} H_k^T (H_k P_{k,k-1} H_k^T + R_{k-1})^{-1} \\
x_k &= x_{k,k} = x_{k,k-1} + J_k (y_k - H_k x_{k,k-1}) \\
P_k &= P_{k,k} = (I - J_k H_k) P_{k,k-1} \\
R_k &= C_{u_k} - H_k P_k H_k^T \\
Q_k &= J_k C_{u_k} J_k^T
\end{aligned}
\tag{8}
$$

where

$$
C_{u_k} = \frac{1}{M} \sum_{i=i_0}^{K} u_i u_i^T \tag{9}
$$

and the residual sequence is $u_k = y_k - H_k x_k$. Here $i_0 = K - M + 1$ is the first epoch of the estimation for a moving window with size M. Empirically the optimal size of the moving window has been determined to be about 25 epochs. $x_{k,k-1}$ and $P_{k,k-1}$ are the predicted state vector and its covariance matrix respectively, and $x_{k,k}$ and $P_{k,k}$ are the estimated ones. J_k is the gain matrix.

A float value of the primary ambiguity N_1 and RTZD can be estimated by using a Kalman filter operating in a recursive manner. For example:

$$
N_1^{1,2} = (x_k^{1,2} - 60\lambda_{77,-60} N_{1,-1}^{1,2}) / 17\lambda_{77,-60} \tag{10}
$$

Then the integer value of N_1 can be fixed conventionally, e.g. using the LAMBDA approach. As a result, the residual vectors V per satellite pair per frequency can be obtained as follows:

$$
v = \begin{bmatrix} \lambda(\phi_{1,2} - N_{1,2}) - \rho_{1,2} \\ \lambda(\phi_{1,3} - N_{1,3}) - \rho_{1,3} \\ \vdots \\ \lambda(\phi_{1,n} - N_{1,n}) - \rho_{1,n} \end{bmatrix} \tag{11}
$$

where subscripts represent the pair of reference stations used for data differencing. As the resuduals are geometrically correlated, a functional model can be applied to Equation (11) in order to compute a pair of coefficients α and β:

$$
v = \begin{bmatrix} X & Y \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \tag{12}
$$

where $\begin{bmatrix} X & Y \end{bmatrix}$ is a vector of baselines between the master reference station and the other reference stations. Note that a pair of coefficients can be obtained per pair of satellites for dual frequencies.

2.2 Network Correction for Virtual Reference Station

Once α and β are obtained from Equation (12), the pseudorange and carrier phase range corrections for a client user within the coverage of the CORS network can be computed on an epoch-by-epoch and satellite-by-satellite basis. The client's approximate position will be used as the location of a virtual reference station (VRS) so that the server can perform carrier phase-based DGPS with respect to the VRS.

Note that one satellite has to be chosen from which other satellite observations can be differenced. This primary or master satellite is normally the satellite with the highest elevation angle relative to the user. The network corrections for the primary satellite for both pseudoranges and carrier phase ranges are:

$$\begin{aligned}
(P_1)_V &= (P_1)_M + \rho_{M,V} + T_{M,V} + (\alpha_1)X_{M,V} + (\beta_1)Y_{M,V} \\
(P_2)_V &= (P_2)_M + \rho_{M,V} + T_{M,V} + (\alpha_2)X_{M,V} + (\beta_2)Y_{M,V} \\
(\lambda_1\phi_1)_V &= (\lambda_1\phi_1)_M + \rho_{M,V} + T_{M,V} - (\alpha_1)X_{M,V} - (\beta_1)Y_{M,V} \\
(\lambda_2\phi_2)_V &= (\lambda_2\phi_2)_M + \rho_{M,V} + T_{M,V} - (\alpha_2)X_{M,V} - (\beta_2)Y_{M,V}
\end{aligned} \tag{13}$$

and the corrections for the other satellites would be:

$$\begin{aligned}
(P_1)_V &= (P_1)_M + \rho_{M,V} + T_{M,V} \\
(P_2)_V &= (P_2)_M + \rho_{M,V} + T_{M,V} \\
(\lambda_1\phi_1)_V &= (\lambda_1\phi_1)_M + \rho_{M,V} + T_{M,V} \\
(\lambda_2\phi_2)_V &= (\lambda_2\phi_2)_M + \rho_{M,V} + T_{M,V}
\end{aligned} \tag{14}$$

The network corrections can be partitioned into dispersive and non-dispersive components. It can be proven that the main contributing factor for the residuals of the ionosphere-free combination $\phi_{77,-60}$ is tropospheric delay, while the ionosphere delay can be derived from the residual of the geometry-free combination. Therefore these residuals can be utilised to generate the dispersive and non-dispersive correction terms. Users can benefit from the availability of such partitioned corrections because they have the flexibility to apply optimal algorithms of their own design.

3. SYSTEM ARCHITECTURE

In server-based RTK, the main server distributes computing resources to the distributed servers. Once a client requests server-based RTK solutions, the main server queries the main database server for the most suitable combination of reference stations that surround the user's position and allocates a distributed server to the client. The main server calculates the network-wide parameters, e.g. orbit errors and atmospheric parameters from the network-wide observations, empirical covariance parameters for Kriging, and so on.

The main server determines whether precise IGS orbits or the network orbits must be used, unless there is a special request from the client. As for server-based network-RTK, the distributed server obtains precise ephemerides from the main database server and generates network corrections by interpolating residuals. Figure 5 shows the system architecture for a distributed-computing scheme such as the one proposed here (Lim & Rizos, 2007).

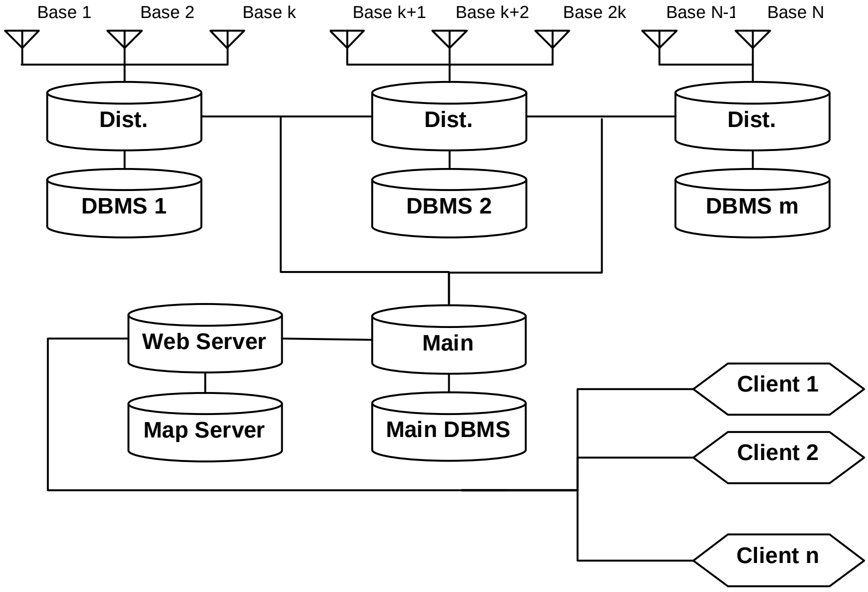

Figure 5: System architecture of a distributed-computing based RTK service.

An NTRIP caster is located at each distributed server and distributes GNSS data coming from NTRIP servers, which themselves access raw measurements from GNSS receivers. A distributed server depicted in Figure 6 is designed to service as many local users as possible. There are two main communication processes running on the distributed server.

1) Reference Stations and Distributed Server

A distributed server communicates with reference stations as an NTRIP client. The incoming binary data must be decoded by RTCM decoders and recorded at a dedicated database in real-time. A GNSS database has been designed and implemented for this purpose. Data records can be extracted at any time (e.g. 1 second later) and converted to RINEX files for post-processing. Network ambiguities and residual vectors are calculated and stored in the distributed DBMS.

2) NTRIP Caster and Distributed Server

The distributed server receives multiple GNSS data streams via TCP/IP from rover receivers as well as from reference stations. Once a client requests server-based RTK by transmitting the raw measurements via NTRIP caster, the distributed server determines the VRS based on the approximate position of the client and computes the coordinates. Each VRS generator can be running inside the distributed server, or anywhere else, because the distributed server and the VRS generator are separated. TCP/IP sockets can be setup to facilitate communications, that is, the VRS generator can be running remotely. This provides a great deal of flexibility and scalability since some clients may want an “exclusive” service for processing their data, others not.

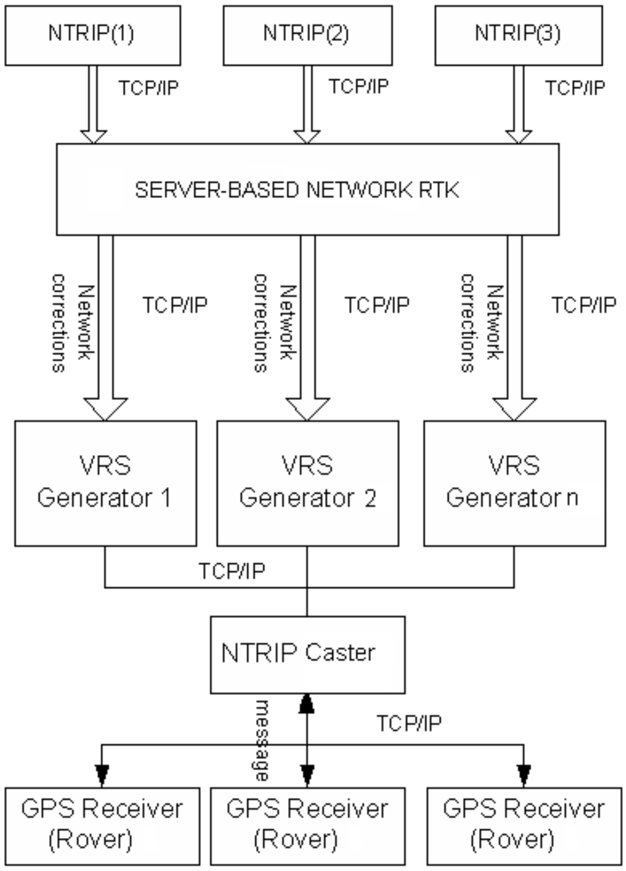

Figure 6: Distributed server for Network-RTK.

4. CONCLUDING REMARKS

In this paper a real-time integer ambiguity resolution procedure that can be used with multiple reference baselines has been presented. The wide-lane ambiguity is first fixed, then the primary ambiguity and relative tropospheric zenith delay are estimated using the ionosphere-free combinations and an adaptive Kalman filter. A server-based RTK system is proposed and its basic design described. The example of computing VRS corrections for server-based network-RTK is presented to illustrate the process. A system architecture for server-based network-RTK is proposed with the consideration of distributed-computing. Distributed-computing allows computers to efficiently communicate with each other and to individually process data. The proposed system design is expected to cope with simultaneous requests from hundreds of clients.

REFERENCES

1. Goad, C.C. (1992). Robust techniques for determining GPS phase ambiguities, Proc. 6th DMA, 1:245-254.
2. Hu, G.R., Khoo, H.S., Goh, P.C., Law, C.L. (2003). Development and assessment of GPS virtual reference stations for RTK positioning, *Journal of Geodesy*, 77, 292–302
3. Lim, S., Rizos, C. (2007). A new framework for server-based and thin-client GNSS

operations for high accuracy applications in surveying and navigation, *20th International Technical Meeting of the Satellite Division of the U.S. Inst. Of Navigation*, 25-28 September, Fort Worth, Texas, 2160-2177.
4. Park, K.D., Nerem, R.S., Schenewerk, M.S, Davis, J.L. (2004). Site-specific characteristics of global IGS and CORS GPS sites, *Journal of Geodesy*, 77, 799-803.
5. Teunissen, P.J.G. (1995). The least-squares ambiguity decorrelation adjustment: A method for fast GPS integer ambiguity estimation, *Journal of Geodesy*, 70, 65-82.
6. Wanninger, L. (2003). Virtual Reference Stations (VRS), *GPS Solutions*, 7, 143-144.
7. Wanninger, L., May, M. (2001). Carrier phase multipath calibration of GPS reference stations, *Journal of the U.S. Inst. of Navigation*, 48(2), 113-124.
8. Wübenna, G., Bagge, A., Seeber, G., Boder, V., Hankemeier, P. (1996). Reducing distance dependent errors for real-time DGPS applications by establishing reference station networks, *9th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation*, Kansas City, Missouri, 17-20 September, 1845–1852.
9. Zhang, J., Lachapelle, G. (2001). Precise estimation of residual tropospheric delays using a regional GPS network for RTK applications, *Journal of Geodesy*, 75 (5-6), 255-266.

BIOGRAPHICAL NOTES

Dr. Samsung Lim is a Senior Lecturer in the School of Surveying and Spatial Information Systems, The University of New South Wales (UNSW), Sydney, Australia. For the past fifteen years his research has focused on the areas of GNSS and GIS. Samsung's research interests are in theoretical problems related to RTK-GPS and applying geo-spatial information technologies to real-world problems. In 2005, Samsung developed an address-based search tool in conjunction with contemporary web-map services such as Google Earth. Samsung received his B.A. and M.A. in Mathematics from Seoul National University and his Ph.D. in Aerospace Engineering and Engineering Mechanics from the University of Texas at Austin.

Prof. Chris Rizos is a graduate of the School of Surveying and Spatial Information Systems, UNSW; obtaining a Bachelor of Surveying in 1975, and a Doctor of Philosophy in 1980. Chris is currently Professor and Head of the School. Chris has been researching the technology and high precision applications of GPS since 1985, and has published over 350 journal and conference papers. He is a Fellow of the Australian Institute of Navigation and a Fellow of the International Association of Geodesy (IAG). He is currently the Vice President of the IAG and a member of the Governing Board of the International GNSS Service.

CONTACTS

Dr. Samsung Lim
School of Surveying and Spatial Information Systems
The University of New South Wales
Sydney, NSW 2052
AUSTRALIA
Tel. + 61-2-9385 4505
Fax + 61-2-9313 7493
Email: s.lim@unsw.edu.au

Prof. Chris Rizos
School of Surveying and Spatial Information Systems
The University of New South Wales
Sydney, NSW 2052,
AUSTRALIA
Tel. + 61-2-9313 4205
Fax + 61-2-9313 7493
Email: c.rizos@unsw.edu.au