http://www.icact.org/journal.asp

ICACT-TACT JOURNAL

Transactions on Advanced Communications Technology

Volume 2 Issue 4, Jul 2013, ISSN: 2288-0003

Editor-in-Chief
Prof. Thomas Byeongnam YOON, PhD.

GIRI Global IT Research Institute

Volume 2, Issue 4

Grid Architecture and Implementation for ALICE Experiment

Jin HUANG*, Pablo SAIZ**, Latchezar BETEV**, Federico CARMINATI**, Costin GRIGORAS**, Steffen SCHREINER***, Jianlin ZHU****

**Research Institute 709, China Shipbuilding Industry Corporation, Wuhan, 430070, China*

*** CERN,1211 Geneva,Switzerland*

**** Technische Universitaet Darmstadt,Germany*

***** South-Central University for Nationalities, Wuhan City, 430074, China*

derick0320@gmail.com, pablo.saiz@cern.ch , latchezar.betev@cern.ch, federico.carminati@cern.ch, costin.grigoras@cern.ch, steffen.schreiner@cern.ch, dczhou.iopp@gmail.com, zhujianlin1120@gmail.com

Abstract **—According to the requirement of the Physics Data Challenge and many years of experiences from the actual operations, we propose the Grid Architecture for the ALICE Experiment. This architecture builds on AliEn2 framework and AliEn2 is the grid framework for the whole system's management and also provides the client interface for accessing the ALICE Grid platform. On the base of AliEn2, we adopt some middlewares to make the platform more flexible, availability and efficiency. At the end, the whole system has been worldwide deployed at around 88 sites for distributed production of raw data and Monte Carlo data, detector simulation, reconstruction and real data analysis. We observe the activities of the platform from different point of view and report the performance of the whole ALICE grid platform. The whole Grid Architecture for ALICE experiment plays an important role in High Energy Physics. It successfully manages the resources distributed all over the world and has a great efficiency system to make all peoples from and in different countries working together. The experiments show the whole system works fine and efficient.**

***Keyword*— Grid, AliEn, Job Management, Transfer Management, Middleware**

Manuscript received May 15, 2013. This work was sponsored by the National Natural Science Foundation of China with Key Grant 11020101060, and the "973" Grant of MOST of China 2013CB837803.

Jin HUANG is with Research Institute 709, China Shipbuilding Industry Corporation, Wuhan, China(Tel:86-13476165997; email:derick0320@gmail.com)

Pablo SAIZ is with CERN, Geneva, Switzerland(email: pablo.saiz@cern.ch)

Latchezar BETEV is with CERN, Geneva, Switzerland(email: latchezar.betev@cern.ch)

Federico CARMINATI is with CERN, Geneva, Switzerland(email: federico.carminati@cern.ch)

Costin GRIGORAS is with CERN, Geneva, Switzerland(email: costin.grigoras@cern.ch)

Steffen SCHREINER is with Technische Universitaet Darmstadt,Germany(email: steffen.schreiner@cern.ch)

Jianlin ZHU is with South-Central University for Nationalities, Wuhan, China(email:zhujianlin1120@gmail.com)

I. INTRODUCTION

Geneva, 30 March 2010, Beams collided at 7 TeV in the LHC at 13:06 CEST, marks the start of the Large Hadron Collider(LHC) research programme. The A Large Ion Collider Experiment(ALICE) is one of the main experiements in LHC, which is dedicated to the study of nucleus-nucleus collisions.

The ALICE experiment produces large amount of data (peta byte data each year), which requires huge computing power to compute and analysis the data, and Peta Byte order of storage to store raw and analysed data. ALICE uses the grid platform[1] to distribute and analyze all raw data and Monte Carlo(MC) data among the 88 sites that participate in the ALICE Collaboration[2].

In this paper, we give the architecture for the whole grid platform and describe the main components in the system for a better understanding the system. At the final, we give the performance analysis of the platform in different views related with services, jobs, transfers and network since the first LHC run.

II. GRID ARCHITECTURE

Grid middleware distributions are often large software artifacts, which include a set of components providing a basic functionality. Such capabilities include (but are not limited to) job management, storage management, authentication and authorization, and resource monitoring. Considering the ALICE experiment:

1) The resources are complex which are distributed all over the world and the reources are increasing every year .

2) The large amount of data (PB) produced by the ALICE experiment every year which is related with Data Management, Transfer Management,etc;

3) Each data analysis job has different requirements such as different softwares and datasets which requires Package Management;

4) The security framework has to be considered to ensure Information Security, Monitoring, Authentication and Authorization;

5) Supporting other Virtual Organizations and middilewares.

With the considerations mensioned above, the Grid Architecture designed for ALICE experiment is described in Figure 1. In Figure 1, there are three main layers for the Grid Architecture:

1) External Software: These are the basic softwares which are used to build the grid platform. The main frameowork is based on AliEn2[3] which uses the perl language. AliEn2 is a framework that has been developed by the Alice Collaboration with the aim of offering to the end user a transparent access to distributed computing and storage resources all over the world. LDAP is used to keep the static information about each site, user, service,etc. The relational database is adopted such as mysql, postgresql, oracle,etc. The softwares required by the Jobs are also supported such as AliRoot, Root, etc.

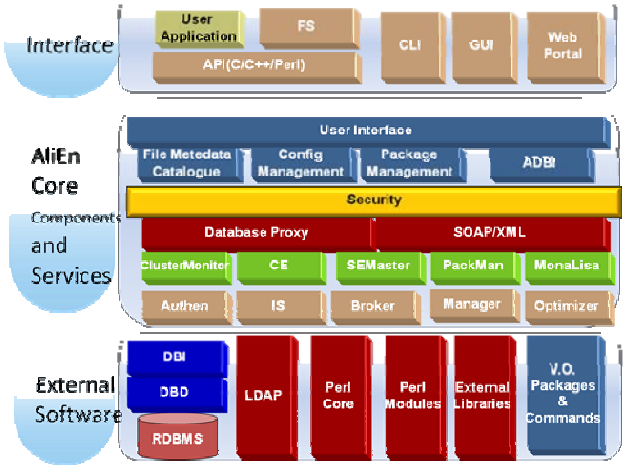


Figure 1. The Grid Architecture for ALICE Experiment

2) AliEn Core Components and Services: The core components of AliEn infrastructure[4] are Web Services, which communicate with each other via XML-based messages using SOAP (Simple Object Access Protocol) protocol. The services at AliEn2 are categorized into two aspects:

- Central services- Authen, IS, Job/Transfer/JobInfo Manager, Job/Transfer Broker,Job/Transfer/Catalogue Optimizer,SE/ PackMan/Messages Master,Logger, API, FTD.
- Site serivces- Cluster Monitor,PackMan, CE, SE, MonaLisa. JobAgent service is running on each worker node in the site.

All the public information about the Central Services and Site Services are published through the LDAP.

3) Interface: The Grid Architecture provides different interfaces for the users to use the resources with different purpose. Such as through the client of AliEn2, users could submit/cancel/resubmit the jobs, add/delete/backup the files. Though the web portal of monitoring system, users could check the system running information in any place any time. With the plugin mechanism, the middlewares could be supported easily.

The required Grid functionality is provied via a combination of the open source softwares, the common middlewares offered by many Grid projects, and the ALICE-specific services from AliEn2.

A. Middlewares

There are some important middlewares that the ALICE Grid Architecture adopts and have already been deployed in the sites all over the world.

1) VOBOX[5]: It focus on the managing the services deployed inside it and softwares for the data analysis, providing renew proxy service for providing long-life proxies. The mechanism in VOBOX is that a proxy renewal service runs for automatic renewal or registering proxies via the MyProxy mechanism to provide long-lived agents and services. the proxy renewal service on the VOBOX uses a setuid root cron job to generate a proxy for the machine, then uses that proxy to authenticate to the MyProxy server in order to renew the registered proxies.

2) CREAM(Computing Resource Execution And Management service)[6,7]: It is a service for job management at the Computing Element(CE) level. to which the CE has interface for supporting and managing different batch system. Moreover, a CE implements additional features with respect to the ones provided by the underlying batch system, such as Grid-enabled user authentication and authorization, accounting, fault tolerance and improved performance and reliability.

3) Xrootd[8,9]: The xrootd system is a high-performance data server. It is useful to build scalable, fault tolerant and high performance storage systems for POSIX-like data access. Unlike NFS, the xrootd protocol includes a number of features like:

- Communication optimizations, which allow clients to accelerate data transfers.
- An extensive fault recovery protocol that allows data transfers to be restarted at an alternate server.
- A comprehensive authentication framework.
- Peer-to-peer elements that allow xrootd servers to cluster together while still providing a uniform name space.

4) CERN Advanced STORage manager(CASTOR)[10,11]: It is a Hierarchical Storage Management system developed at CERN used to store physics production files and user files. Files can be stored, listed, retrieved and accessed in CASTOR using command line tools or applications built on top of the different data transfer protocols like RFIO (Remote File IO), ROOT libraries, GridFTP and XROOTD. CASTOR manages disk cache(s) and the data on tertiary storage or tapes. Xrootd is one of the protocols supported for data transfer and access.

B. Central Services and Site Services

The LHC experiment has four experiments (ALICE,ATLAS,CMS and LHCb) and each experiment is one Virtual Organization(VO). One VO only has one set of Central services and could have many site services.

A typical deployment example in one VO is given in Figure 2 to show how the platform works.

The Central services along with LDAP server, File Catalogue, TaskQueue and TransferQueue take responsible for the management of the whole system and they are deployed at CERN. In the section, we will discuss detailed relationship between these services.

In Figure 2, to manage the local resources in one site, the following components are needed:

1) Site Services (Cluster Monitor, PackMan, CE, MonALISA) for communicating with the central services to ask for jobs to execute;

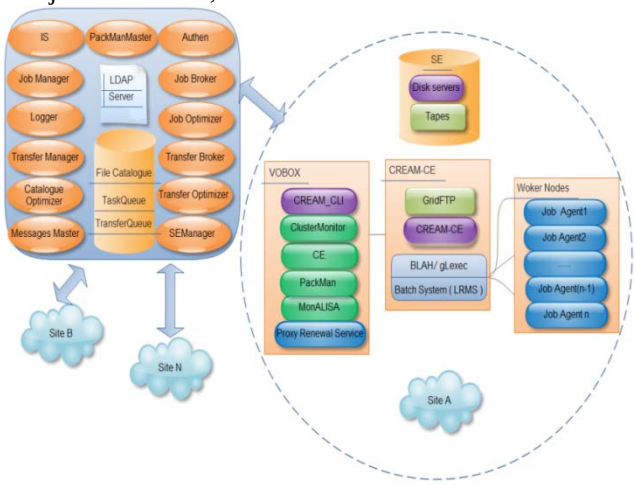


Figure 2. Central services and Site services in a VO

2) VO-BOX for providing a reliable mechanism to the site services to accomplish various tasks(Start/Stop services deployed at each site, renew proxies, installed software management, etc).

3) CREAM-CE for getting the JobAgents from CE and distributing them to worker nodes.

4) JobAgent service running on each worker node for doing some tasks to make sure that the analysis jobs could be executed automatically and successfully, monitoring the status of the analysis jobs, reporting the information about the running jobs and asking more jobs to execute when the job is finished.

5) The Storage Element is used for managing the file-based data. Each site provides certain storage space for the experiment and storage could be tape and disk. In Figure 2, the xrootd protocol is chosen as the transfer protocol for data transfer and data access.

III. IMPLEMENTATION

A. Jon Management

In the system, Job Management is mainly related with Job Manager, Job Broker, Job Optimizer, JobInfo Manger , Task Queue and Authen from Central Services, CE, Cluster Monitor, JobAgents in one site.

The whole workload management system is based on a central job queue which holds all tasks that have to be executed. The Job Broker model is based on pilot jobs: the system submits generic pilots to the computing centres batch gateways, and the assignment of a real job is done only when the pilot wakes up on the worker node. The model facilitates a flexible fair share user job distribution.

A Computing Element(CE) is some set of computing resources localized at a site. A CE includes a Grid Gate(GG), which acts as a generic interface to the cluster; a Local Resource Management System(LRMS), and the cluster itself, a collection of Worker Nodes(WNs) where the nodes the jobs run. The Grid Gate in ALICE is CREAM-CE, which is responsible for accepting jobs and dispatching them for execution on the WNs via the LRMS. The LRMS types are OpenPBS/PBSPro,LSF,Maui/Torque,BQS and Condor.

The jobs in ALICE experiment are described by the Job Description Language(JDL). The users submit the jobs to the Central services , the related services processes the JDL, inserted into the Task Queue and distrbuted the jobs to the proper sites. Figure 3 shows the ***Job Flow Model*** of the platform. The main components in the ***Job Flow Model*** is as follows:

1) TaskQueue is the database with all the Jobs that have to be done. The heart of the job management system is the central TaskQueue which calculates the priorities of the jobs. SITE_CE is the database with all the jobs detail information running in each site.

2) Job Manager service is responsible for accepting jobs submitted from the users and inserting jobs into TaskQueue, checking the job quota of the user, managing each status of JobAgent and each file generated by the jobs.

3) Job Broker service is in charge of distributing the jobs to each site matching the requirement of the jobs. If the TaskQueue is the heart of the job model, the Job Broker is the mind. The Job Broker is responsible for deviding which job will be executed on site. The decision will be made depending on the data that the job has to analyze, the software packages, time needed by the job and any specific user requirements(name of the site, partition, disk space, etc). The coversation between Job Broker and CE use pull model, and it means that CEs periodically ask the Job Broker whether it has jobs to be executed. If the JDL of the CE matches the JDL of Jobs, Job Broker will send the JDLs of the jobs to that CE. Otherwise, the CE will sleep for a while and then ask the Job Broker again.

4) Job Optimizer service does JDL parameter optimization for all the jobs waiting to be executed. According to the information gathered from the whole system, the Job Optimizer could change the content of JDL and make the jobs easier be matched with the JDL of CE. Along the same lines,

one can also implement policy monitors to enforce VO policies by altering job priorities.the Broker can modify the job's JDL entry by adding or elaborating requirements based on the detailed information it can get from the system like the exact location of the dataset and replicas,client and service capabilities. The job execution model AliEN uses a pull model. In a pull implementation,the Broker does not need to maintain the status of the system.

From the Figure 3, we could see that after the CE in one site get JDL of jobs from the Central services, it will cooperate with CREAM-CE to send the jobs to the worker nodes for execution.

The execution of jobs is based on central services, and the site services act as the interface to the local resources. The site services announce their capabilities, and based on that, the central services will decide which job can be assigned there. Acorrding to Figure 4, the whole status of job execution could be described in four stages:

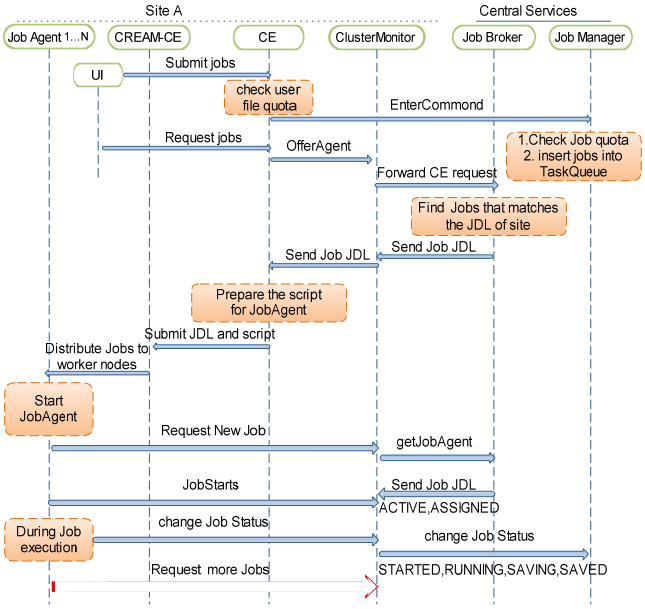


Figure 3. Job Flow Model

1) Submission: The user submits jobs to the CE service. The CE service gets the JDL for this job, then reports this job to the Job Manager Service. The Job Manager Service firstly checks whether the user has enough job quota to execute the job, then does some extra preparement work for this job, and update the job status to "INSERTING".

2) Optimizer: The Job Optimizer Service is used to optimize the jobs with the "INSERTING" status. The Job Optimizer Service firstly inits the execution environment in AliEn2 client shell for the job, prepares the job token, inserts the job into TaskQueue and updates the job status to "WAITING".

3) Broker/Agent: The JobAgent Service asks the ClusterMonitor Service for jobs to execute. The ClusterMonitor sends messages to the Job Broker Services whether it has the suitable jobs for this site. The Job Broker Service checks the queue status of this site whether it is open or not, finds the suitable jobs which match this site and sends them to the ClusterMonitor Services, and updates the job status to "ASSIGNED". After receiving the jobs from ClusterMonitor, The JobAgent starts to run the jobs. According the jobs' JDL, it creates directories , starts the monitoring process to report the status to ClusterMontior which will send the message to the Job Manager Service. During the execution of the job, the job status could be "STARTED", "RUNNING", "SAVING", "SAVED". At the same time, the JobAgent also includes starting monitor, sending JobAgent status, executing the job, checking process, executing ValidationScript, registering logs, etc. The JobAgent status could be "JOB_STARTED", "RUNNING_JOB", "SAVED".

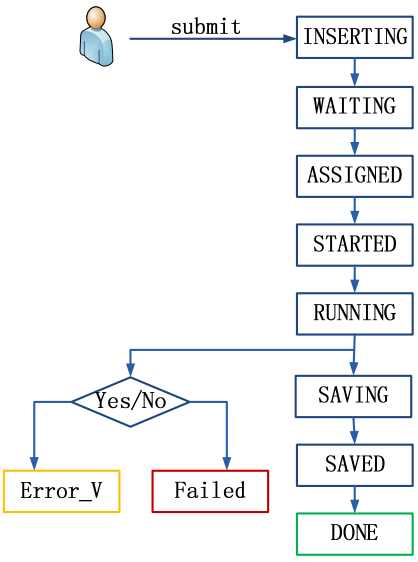


Figure 4. General status job execution diagram

4) Output registration: The Job Optimizer Service finds the jobs marked with "SAVED" and "SAVED_WARN" from TaskQueue, then creates the output information in the user's AliEn client, registers the files in the catalogue, and updates the job stauts to "DONE".

B. Transfer Management

With the consideration of safety of the data, resource usage, ALICE group uses different strategies to deal with data taken from the detectors, data from the users, and data from the jobs. In ALICE experiment, one copy of the raw data from the experiment will be stored at CERN(Tier 0) and a second copy will be distributed among the external Tier 1 centres outside of CERN, thus providing a natural backup. Reconstruction to the Event Summary Data(ESD) level will be shared by the Tier1 centres, with the CERN Tier 0 responsible for the first reconstruction pass. Subsequent data reduction to the Analysis Object Data(AOD) level, all Tier 1 and 2 centres will

participate analysis and Monte Carlo production. The Tier 1s will perform reconstruction and scheduled ananlysis, while the Tier 2s will perform Monte Carlo production and end-user analysis. For all procedures and data analysis mensioned above, the system faces a large mount of data transfer and needs an efficiency Transfer Management to handle all the situations.

The Transfer Management in ALICE experiment includes file catalogue, Transfer Manager, Transfer Broker, Transfer Optimizer, Transfer Queue, FTD and other Storage Management middlewares(Xrootd, CASTOR, etc). The grid platform provides the basic command to store, copy and delete files. Futher more, ALICE also provides functions dealing with the multi-copies of the data, the data transfer between different data server and the data access to the files.

The files are identified by a Logical File Name(LFN) in the grid platform. The files aslo are identified by a Globally Unique Identifier(GUID),which is a fixed-format string generated by the middleware and guaranteed to be absolutely unique. When the user requests a transfer, the LFN, the destination Storage Element(SE) and the type of transfer have to be specified. The JDL syntax for transfer jobs has the following fields: LFN of the file to be replicated; SE where the file has to be replicated; size of the file; Type of replication (Cache,mirror or masterCopy); Requirements of the transfer. The meaning is as follows:

1) Cache: the transferred file will not be registered in the catalogue.

2) Mirror: the new PFN will be inserted in the catalogue,and any user from the same site will get this new PFN instead of the original.

3) masterCopy: the PFN will be registered as the master PFN, and the previous PFN will be mirrors.

In Figure 5, The main components in the ***Transfer Flow Model*** is as follows:

1) A database(TransferQueue) with all the transfers that have to be done; the file catalogue stores the connections between the LFN and pointers to any replicas.

2) Transfer Broker service keeps a description in JDL for each FTD. The description includes name of the FTD; SE close to the FTD; free space in the cache. Each FTD sends its JDL to the Transfer Broker, and the Broker tries to match it against the list of transfers that have to be done. If there is an action that the FTD could do, the Broker sends a JDL with the description of what the FTD is supposed to do. If there was no action, the FTD will sleep for a while and request another transfer when it wakes up.

3) Transfer Manager service is responsible for inserting transfer in the system, changing their status and measuring the time each transfer spent in each state.

4) Transfer Optimizer service is responsible for taking all the new requested transfer, and specifying the PFN and SE that currently has that LFN.

FTD supports many protocols,such as xrd3cp,xrdcp,bbftp,cp,GridFTP,GSIFTP,etc. Most of the sites support xrdcp and xrd3cp. Xrd3cp and xrdcp are able to transfer data files at high speed by using a large number of parrallel TCP connections.

According Figure 6, the transfer status could be categories into four stages.

1) Mirror: When the system receives the "mirror" command, firstly checks the permission of the user, checks the path whether the user has permission to use, gets the information about where the source data is kept, reports the transfer to the Transfer Management Service. The Transfer Management Service registers this transfer information into the Transfer Database.

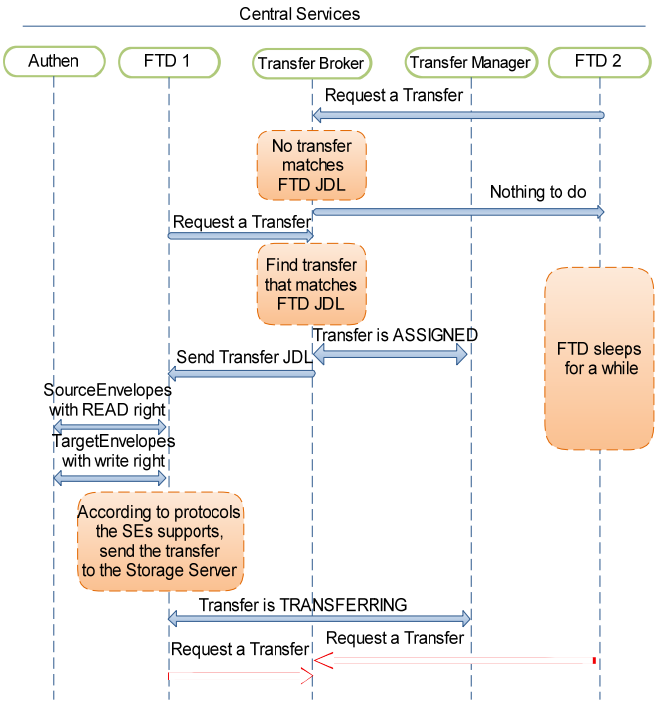


Figure 5. Transfer Flow Model

2) Optimizer: The Transfer Optimizer Service changes the status of this transfer into "INSERTING". The Transfer Optimizer Service got the information such as SE, guid, LFN,etc, then finds the common protocols for this transfer, and updates the status of transfer into "WAITING".

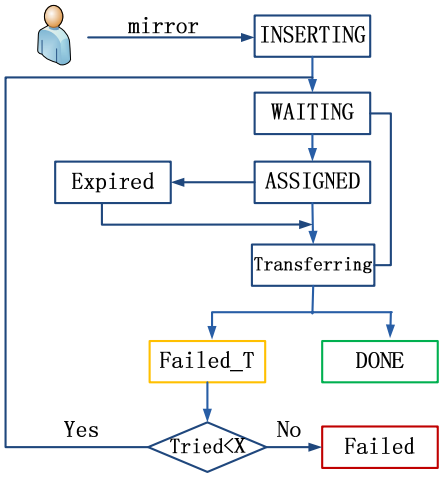


Figure 6. General Status Transfer diagram

3) Broker: The Service FTD requests the transfer from the Broker Manager Service, and updates the status of transfer into "ASSIGNED".

4) FTD: The Service FTD got the transfer JDL , asks the Transfer Manager Service changes the status of transfer to "TRANSFERRING". The Service FTD starts to prepare the envelope information for transferring the files. First the FTD gets user and access information in both source SE and target SE from the Authen Service. The FTD starts the transfer with the setting protocol and asks the Transfer Manager Service to change the status tranfer to "DONE" when the transfer is finished.

C. Packages Management

AliEn provides a service for automatic installation of the software packages required by the job. The installation, upgrade, configuration and removal of application software packages are done by the AliEn Package Manager (PackMan) service. This service is a helper service that automates the process of installing, upgrading, configuring, and remove software packages from a shared area (software cache) on a Grid site. The Package Manager Service does not manage the installation of middleware software; it manage the installation of middleware software; it manages the packages that are common for all users of a VO and possibly packages provides by individual users.

The packages are installed on demand, when requested by the Job Agent running on a worker node or upon an explicit request by the VO Software Manager. The Package Manager checks if the requested package is already installed in the cache and if that is not the case, it proceeds with the installation. The right version of the package is downloaded from the File Catalogue and installed in the directory specified as a service configuration parameter. It will also install all the dependencies required by the package. The Package Manager returns a string with a command (a shell script) that the client has to execute to configure the package and all its dependencies. This script performs actions like setting environment variables. Figure 7 shows which commands use could use to install the packages in the client interface of AliEn2.

The package installation can be triggered by a process running on the worker node or by a person with appropriate privileges (identified by the certificate subject). In both cases the requestor obtains the lease for the package for a specified time. The Package Manager manages the local disk cache and will clear the disk space if it needs the disk space to install newer packages but it will not remove the packages for which someone holds the lease. The maximum lease time for the packages is a configurable parameter. While any user or process can list already installed packages, only the VO administrator can remove a package from the local cache regardless of its current lease status and in that case the currently running jobs requiring that package might fail. Removing a package does not remove the packages that depend on it. If any of those packages are used, the removed package dependency will be automatically installed again.

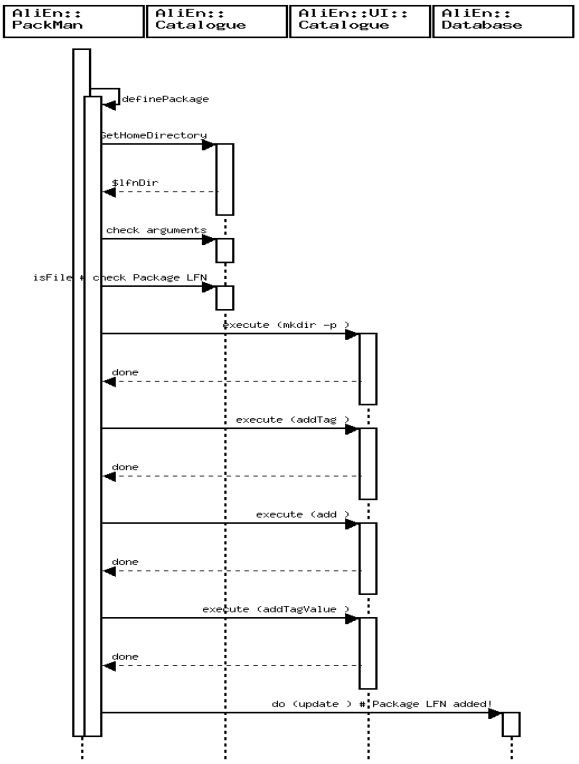


Figure 7. The process of PackMan

D. Security Framework

Considering that the resources for the grid platform in ALICE experiment distribute all over the world, the Security Infrastructure for AliEn2 can be divided into five layers. General overview of these levels is described in Figure 8.

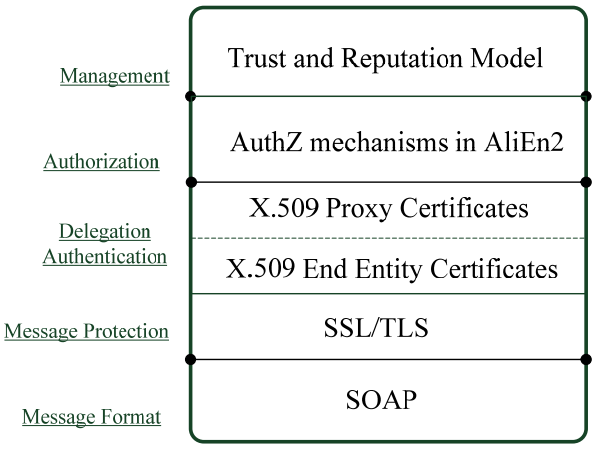


Figure 8. The Security Framework for AliEn2

1) Layer 1: Message Format: the message exchanged between different services is based on Simple Object Access Protocol;

2) Layer 2: Message protection: the protection for the communication channel for AliEn2 adopts SSL/TLS protocol. SSL/TLS protocol ensures the transport level security. It provides for both integrity protection and privacy (via encryption);

3) Layer 3: Delegation/Authentication: The credentials in AliEn2 follows X.509 specification. X.509 EECs are for global authentication. They provide each entity with a unique identifier and a method to assert that identifier to another party through the use of an asymmetric key pair bound to the identifier by the certificate;X.509 proxy certificates are extension to X.509 certificates. GSI proxy certificates and VOMS proxy certificates as a means of authenticating user and delegating authority from users to remote servers. They could bridge disparate security mechanisms, basic support for delegation, policy distribution, single sign-on;

4) Layer 4: The Authorization uses AuthZ mechanisms in AliEn2. AuthZ mechanisms are based on the File Catalogue and Quota to authorize the uses to access a certain set of resources.

5) Layer 5: Management: there could be many different and effective managements integrated into this layer, such as Trust and Reputation management could be designed for improving the efficiency of the AliEn2. Monitor system is to monitor the resource usage. Trust management system to create, negotiate, and manage trust between other systems or "strangers".

The detail implementation of the information security and Authentication could see the paper[12]. The authorization for file catalogues, files transfer could check paper[13].

IV. EXPERIMENTAL RESULTS

For ALICE experiment, we use MonALISA to monitor the whole grid platform. MonALISA[14] is a globally scalable framework of services to monitor and help manage and optimize the operational performance of Grids, networks and running applications in real-time. The information starts from March 20,2010. The information provided by MonALISA could be seen from the statistics listed as follows.

Figure 9 is the entrance of MonALISA monitoring system for ALICE experiment and it shows the MonALISA Repository for ALICE experiment and the MonALISA collect more than 1700 metrics.

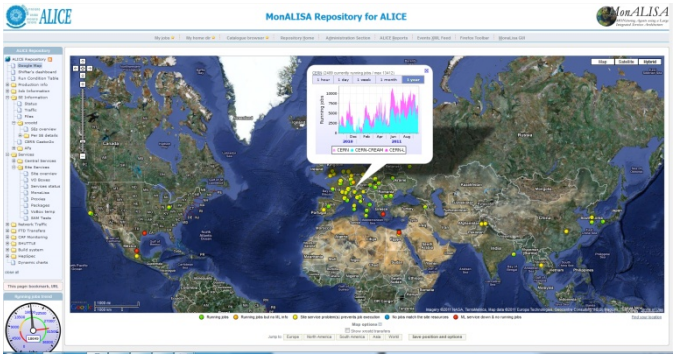


Figure 9. MonALISA Repository for ALICE

Figure 10 is the number of running jobs every day in each site. The peak number of jobs which are running at the same time in the grid system is 32538.

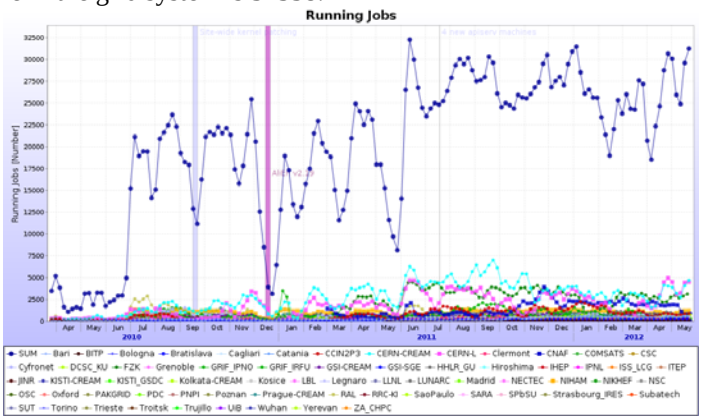


Figure 10. The total number of running jobs per day

Figure 11 describes the RAW data collected from the detectors of ALICE experiment. This is the real data which was recorded during the beam-beam collision and the maximum size of the RAW data is close to 6PB. From the Figure 11, the update of the softwares has been done during the closure of detectors.

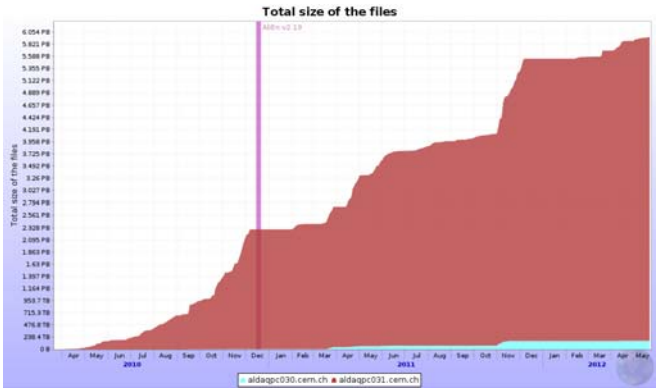


Figure 11. Raw data taken from detectors of ALICE

Figure 12 shows the network traffic of some SEs. The maximum incoming traffic is 87.51 MB/s and the maximum outgoing traffic is 2.341 MB/s. The central services will arrage the transfer jobs according to the traffic size of each SE which could avoid the traffic bottleneck.

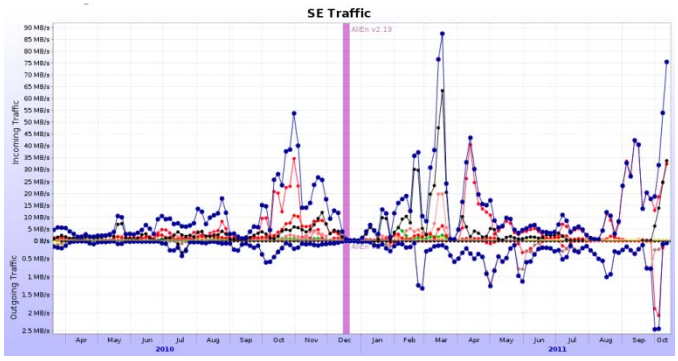


Figure 12. Incoming Traffic and Outgoing Traffic of SEs

There are 8 tape storage elements(ALICE::CCIN2P3::TAPE, ALICE::CERN::CASTOR2, ALICE::CERN::TOALICE, ALICE::CNAF::TAPE,ALICE::FZK::TAPE, ALICE::NDGF::DCACHE_TAPE,ALICE::RAL::TAPE,ALICE::SARA::DCACHE_TAPE) which are used for RAW replication. Figure 13 shows the transfer rate for each tape storage element and the maximum rate is 220.5 MB/s.

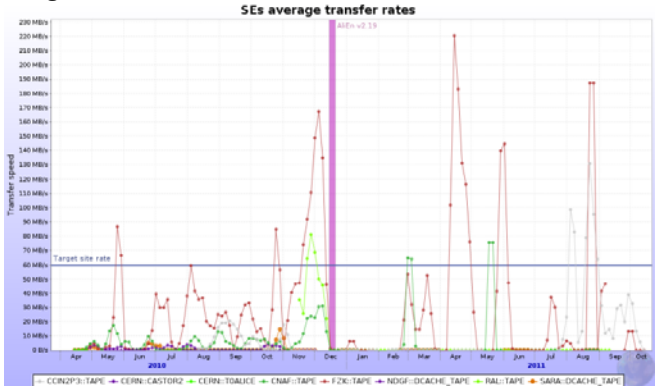


Figure 13. The transfer rates of tape storage elements

There are 54 disk storage elements and 8 tape storages in grid plarform of the ALICE experiment. In Table 1, total size of all the storage elements are 43.67PB and 18.344 PB is used. There are 187,720,276 files saved in the SE.

TABLE 1. DIFFERENT TYPE OF STORAGE ELEMENTS

Storage Type	Total size	Used size	Number of files
Disk storage	12.87 PB	6.454 PB	165,434,307
Tape storage	30.8 PB	11.89 PB	22,285,969

There are many other functions to provide information to site administrators to understand the system and tuning their system to get a better performance. Such as there is a Bandwidth test function in the network traffic, choosing one site, you could get the related information (i.e. site name, speed, hops, RTT, etc) in traffic in and out views with other sites ranked by speed.

Until now, about 88 sites join the ALICE grid platform and contributes the resource to the experiment. Around 43 high performance servers provide central services. We have about 54 disk storage elements and 8 tape storages elements the size of which are around 43.67 PB.

V. CONCLUSION

The whole Grid Architecture for ALICE experiment plays an important role in High Energy Physics. It successfully manages the resources distributed all over the world and has a great efficiency system to make all peoples from and in different countries working together.

A lot of experiences are got from operating the grid system. As the increasing number of users and resources, there are many works on-going:

- Developing interface to more middlewares.
- Developing the Java version AliEn2 platform[15].
- Developing Trust Model for reliability and efficiency.
- Optimizing the whole grid platform.

ACKNOWLEDGMENT

I would like to thank the entire members of ALICE experiments and CERN who do the contribution to the development and deployment of the Whole system. Thanks to the ALICE offline group for giving me the chance to join the development of AliEn2 project and helping me in many areas.

The work is supported by the NSFC (Key Grant 11020101060, IRG11221504 and 11005044), the CCNU11Z01001(Key Grant), the Program of Introducing Talents of Discipline to Universities of China (B08033) .

REFERENCES

[1] The ALICE Collaboration, ALICE: Technical Design Report of the Computing, Printed at CERN June 2005, ISBN 92-9083-247-9.

[2] (2011)The ALICE Collaboration website. [Online]. Available: http://aliweb.cern.ch/
[3] (2011)The AliEn2 website. [Online]. Available: http://alien2.cern.ch/
[4] S Bagnasco et al, "AliEn: ALICE environment on the GRID", Journal of Physics: Conference Series, Volume 119,Part 6, 062012, 2008
[5] P.M. Lorenzo, IT/GS, CERN,"VOBOXes and LCG Services" in Alice T1/T2 Tutorial for Site Admins,CERN,Geneva, 26-27 May 2009.
[6] Cristina Aiftimier, etc, "Design and implementation of the gLite CREAM job management service", Future Generation Computer Systems, Volume 26, Issue 4, April 2010, Pages 654-667.
[7] (2011)The CREAM website. [Online]. Available: http://grid.pd.infn.it/cream/
[8] Dorigo A, Elmer P, Furano F and Hanushevsky A, "XRootD/TXNetfile: a highly scalable architecture for data access in the ROOT environment", Proceedings of the 4th WSEAS International Conference on Telecommunications and Informatics (World Scientific and Engineering Academy and Society (WSEAS)), Wisconsin, USA, 2005.
[9] (2011) The Xrootd website. [Online]. Available: http://xrootd.slac.stanford.edu/
[10] J. P. Baud, B. Couturier, C. Curran, J. D. Durand, E. Knezo, S. Occhetti, O. Barring, "CASTOR status and evolution", Computing in High Energy and Nuclear Physics, 24-28 March 2003, La Jolla, California.
[11] (2011) The CASTOR website. [Online]. Available: http://castor.web.cern.ch/
[12] J. Zhu, P. Saiz, F. Carminati, L. Betev, D. Zhou, P. M. Lorenzo, A. G. Grigoras, C. Grigoras, F. Furano, S. Schreiner, O. V. Datskova, S. S. Banerjee and G. Zhang, Enhancing the AliEn Web Service Authentication, 2011 J. Phys.: Conf. Ser. 331 062048.
[13] S. Schreiner, S. Bagnasco, S. S. Banerjee, L. Betev, F. Carminati, O. V. Datskova, F. Furano, A. G. Grigoras, C. Grigoras, P. M. Lorenzo, A. J. Peters, P. Saiz and J. Zhu, Securing the AliEn File Catalogue – Enforcing authorization with accountable file operations, 2011 J. Phys.: Conf. Ser. 331 062044.
[14] (2011)The MonALISA website. [Online]. Available: http://alimonitor.cern.ch
[15] (2011)The JAliEn website. [Online]. Available: http://jalien.cern.ch/

Dr. Jin Huang, obtained his Ph.D degree in Software Engineering from Huazhong University of Science and Technology, China. He obtained his BSc from Huazhong University of Science and Technology, MSc from Huazhong University of Science and Technology, China. He got a University Scholarship for one year study at CERN from 2009 – 2011. He was a visiting student of HKUST from September, 2008 to March, 2009.

Mr. Pablo Saiz. He is a fellow of CERN. His research area is grid systems. The published papers include "AliEn: ALICE Environment on the Grid", "AliEn resources brokers", "Alienfs-a linux file system for the alien grid services", "Experiment Dashboard: the monitoring system for the LHC experiments".

Dr. Latchezar Betev. He is working in the offline team of the ALICE collaboration at CERN and is responsible for the operation of the grid infrastructure of the experiment.His main interests include large-scale distributed computing and monitoring and control of remote systems. The published papers include "Monitoring and control of large systems with MonALISA", "Monitoring, accounting and automated decision support for the alice experiment based on the MonALISA framework".

Dr. Federico Carminati. He obtained the Italian doctors degree in High Energy Physics at the University of Pavia in 1981. After working as an experimental physicist at CERN, Los Alamos and CalTech, he was hired at CERN were he has been responsible for the development and support of the CERN Program Library and the GEANT3 detector simulation MonteCarlo. From 1994 to 1998 he has participated in the design of the Energy Amplifier under the guidance of Prof. C.Rubbia (1984 Nobel Physics Laureate) in the development of innovative MonteCarlo techniques for the simulation of accelerator driven fission machines, and of the related fuel cycle. In January 1998 he has joined the ALICE collaboration at LHC assuming the leadership of the ALICE software and computing project. Since January 2001 he is holding the position of Work Package Manager in the European DataGRID project. He is responsible for the High Energy Physics Application Work Package whose aim is to deploy large scale distributed HEP applications using the GRID technology.

Dr. Costin Grigoras. He obtained his bachelor and master degree in computer science from Politehnica University of Bucharest, Romania in 2004. He obtained his PhD degree in computer science from Politehnica University of Bucharest, Romania in 2009. He is a fellow at CERN from March, 2006 to present. He is currently focusing on distributed systems monitoring, automated decision taking and optimizing databases and code to handle high volumes of data. He is also interested in parallel processing, multicore-aware applications and new technologies in general.

Mr. Steffen Schreiner. He is a PhD *ca*ndidate in Technische Universitaet Darmstadt,Germany. His research area is security of the distributed system. The published papers include "Enhancing the AliEn Web Service Authentication", "Securing the AliEn File Catalogue", "A tool for optimization of the production and user analysis on the Grid" at CHEP2011, "A security architecture for the ALICE grid services" at ISGC2012, "A Mediated Definite Delegation Model alowing for Certified Grid Job Submission", etc.

Jianlin Zhu, Yancheng City, Jiangsu Province, China. November 20, 1981. She obtained her Bachelor degree in Computer Science from School of Computer Science and Technology, Huahzong University of Science and Technology, Wuhan City, China in 2004. She obtained her Master degree in Computer Architecture from School of Computer Science and Technology, Huahzong University of Science and Technology, Wuhan City, China in 2007. She is currently a Ph.D student in College of Physical Science and Technology, Central China Normal University, Wuhan City, China. Her research work focuses on grid platform, security of distributed system, program analysis.

Performances Analysis of Algebraic Space Time Code under Correlated and Uncorrelated Channels

Ines BEN HASSINE *, Ridha BOUALLEGUE *

* SUPCOM, InnovCOM Laboratory, National Engineering School of Tunis

Tunis, Tunisia

ines.benhassine@yahoo.fr, ridha.bouallegue@supcom.rnu.tn

***Abstract*— With their very Algebraic-construction based on Quaternionic algebra, Algebraic Space Time Codes (ASTC), called the Golden codes, have a full rate, full diversity and non-vanishing constant minimum determinant for increasing spectral efficiency. They have also uniform average transmitted energy per antenna and good shaping, readily lend themselves to high data rate situations. In this paper, we first analyze the performances of the ASTC codes in correlated Rayleigh channel. We consider a coherent demodulator using different decoding schemes and we analyze the Bit Error Rate (BER). In order to increase the spectral efficiency and to maximize the coding gain, ASTC have been proposed for MIMO flat fading channels. To deal with the frequency selectivity, we use the OFDM modulation. So we analyze the performances of an ASTC-MIMO-OFDM system in terms of BER. Finally, we investigate the impact of spatial correlation on the ASTC code design in terms of BER and capacity.**

***Keywords*— ASTC code, OFDM, MIMO, Rayleigh Channel, spatial correlation, capacity, Bit Error Rate**

I. INTRODUCTION

Orthogonal frequency division multiplexing (OFDM) has been attracting considerable attention because of its robustness against frequency-selective fading [1]. OFDM system has been adopted as a standard for digital audio broadcasting, digital video broadcasting, and broad-band indoor wireless systems thanks to his efficiency combat inter-symbol interference (ISI). In fact, it is considered as an effective method for high-rate communication systems [2]. On the other hand, information theory indicates that a multi-input–multi-output (MIMO) system is able to support enormous capacities [3], provided the multipath scattering of a wireless channel is exploited with appropriate space–time signal-processing techniques. However, the MIMO system requires a complicated channel-equalization technique in a frequency-selective broad-band channel, in order to eliminate the ISI. The use of an OFDM technique for MIMO systems would be desirable to alleviate this problem. Recent studies have shown that high-performance transmission can be provided by combining the OFDM technique with a MIMO system [4].

Manuscript received February 13, 2013

I. BEN HASSINE and R. BOUALLEUE are with National Engineering School of Tunis, Innovation of COMmunicant and COoperative Mobiles Laboratory, Tunis, Tunisia. (corresponding author, I. BEN HASSINE; e-mail: ines.benhassine@yahoo.fr

By providing a temporal and a spatial multiplexing modulation, the space-time codes are used to improve MIMO performances. The Alamouti code [5] and the Golden code [6] represent the most known and used Space-Time Block codes (STBCs). The Golden code, which has been proposed in 2004 for 2*2 MIMO system, is a full-rate and full-diversity space-time code that has a maximal coding gain. Thanks to its algebraic construction, it will be shown in this paper that the ASTC codes outperforms the Alamouti codes in flat fading channels. In a frequency-selective channel, the ASTC codes lose their proprieties due the inter-symbol interference (ISI). The orthogonal frequency division multiplexing (OFDM) modulation can overcome this problem.

MIMO is used for transmitting multiple data streams or increasing the reliability (in terms of BER) assuming that the propagation channels between each pair of transmit and receive antennas are statistically independent and identically distributed. Or in practice, the channels between different antennas are often correlated which is called spatial correlation, interpreted as a correlation between a signal's spatial direction and the average received signal gain. The spatially correlated MIMO channels can substantially reduce system performances. So we propose to study the impact of spatial correlation on the ASTC code design performances.

In this work, we first propose a coded ASTC system in flat fading channel. Then, we analyze the ASTC-MIMO-OFDM system in Rayleigh selective-channel. We use a data aided channel estimation method based on the pilot symbol insertion in the detector to deduce the channel transfer function. Finally, we analyze ASTC-MIMO-OFDM system performances under spatially correlated channel.

This paper is organized as follows. In section II, we present the main criteria of the proposed coded ASTC chain. The third section focuses on ASTC-MIMO-OFDM system. In section IV, we analyze the ASTC-MIMO-OFDM system performances under spatially correlated channel. In section V, we present simulation results. Finally, a conclusion is given in section VI.

II. CODED ASTC SYSTEM

In this section we first propose the system model of a coded ASTC chain, then the channel model is introduced, and finally the decoder structure is described.

A. System Model

At first, equally probable random numbers (the data to be transmitted) are created in Matlab and mapped onto a 4-QAM constellation such that the possible symbol values are 1+ j, 1- j, -1+ j, and -1- j. Each information sequence S is encoded by the ASTC encoder.

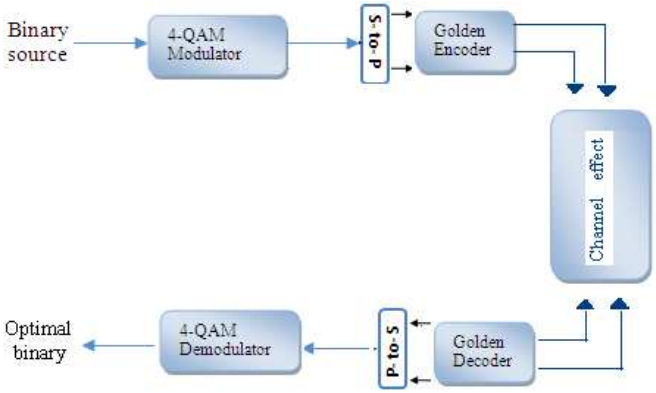


Fig. 1. Block diagram of Coded ASTC System

The algebraic construction yields code-words of the Golden code of the form

$$C = \frac{1}{\sqrt{5}} \cdot \begin{pmatrix} \alpha(a+\theta b) & \alpha(c+\theta d) \\ \bar{\alpha}(c+\bar{\theta} d) & \bar{\alpha}(a+\bar{\theta} b) \end{pmatrix} \quad (1)$$

where $\theta = \frac{1+\sqrt{5}}{2}$, $\bar{\theta} = \frac{1-\sqrt{5}}{2}$, $\alpha = 1+i-i\theta$, $\bar{\alpha} = 1+i-i\bar{\theta}$, a, b ,c and d are the 4-QAM modulated symbols.

In MIMO systems, the general transmission model is

$$Y = HX + W \quad (2)$$

where X is the transmitted codeword, H is the channel matrix and W is the i. i. d. Gaussian noise matrix.

To have full-rate square codes using QAM constellation, we consider square (2*2) linear dispersion. We can express the code word X as the result of multiplication of each four consecutive symbols of information sequence S by the matrix Φ_t.

$$\Phi_t = \begin{pmatrix} \alpha & \alpha\theta & 0 & 0 \\ 0 & 0 & \overline{i\alpha} & \overline{i\alpha\theta} \\ 0 & 0 & \alpha & \alpha\theta \\ \bar{\alpha} & \overline{\alpha\theta} & 0 & 0 \end{pmatrix} \quad (3)$$

So at time (t,t+1), we can express the vector X_t as follow, where the first two lines are transmitted over antenna 1, and the rest two ones are transmitted over antenna 2.

$$X_t = \begin{pmatrix} (\alpha(a+\theta b))_{(t,1)} \\ \bar{\alpha}(c+\bar{\theta} d))_{(t+1,1)} \\ (\alpha(c+\theta d))_{(t,2)} \\ (\bar{\alpha}(a+\bar{\theta} b))_{(t+1,2)} \end{pmatrix} \quad (4)$$

B. Channel Model

In this work, we suppose that the encoded signal is transmitted over a non selective correlated Rayleigh fading channel. We consider here the Clarke channel model. The resulting sequence X will be transmitted over a non selective channel H. We can express the elementary matrix H_t at time (t,t+1) as:

$$H_t = \begin{pmatrix} h_t^{11} & h_{t+1}^{21} & 0 & 0 \\ h_t^{12} & h_{t+1}^{22} & 0 & 0 \\ 0 & 0 & h_t^{11} & h_{t+1}^{21} \\ 0 & 0 & h_t^{12} & h_{t+1}^{22} \end{pmatrix} \quad (5)$$

We note here that the encoder can transmit 4 symbols on each antenna at the same time, whereas the Alamouti [5] encoder can only code 2 symbols at a time.

C. Decoder structure

We will consider two structures of decoders: We decode the received signal using Brute Force ML Decoding and Sphere Decoding Algorithm.

1) ML decoding

The best performance is given by the brute force ML decoder which searches for the matrix X which minimizes the overall noise power. i.e. an ML decoder computes an estimate of the transmitted matrix as

$$\hat{X} = \arg\min_{X} \|Y - HX\|^2 \quad (6)$$

But the ML decoder has a very high complexity in MIMO channels. To lower the complexity, a new type of decoding method called sphere decoding can be used. The sphere decoding algorithm has near ML performance with reasonably low complexity [7].

2) The Sphere Decoding Algorithm

The principle of sphere decoding algorithm is to search the closest constellation point to the received signal with in a sphere of some initial radius. If a point is found and if the distance between the centre and the point is less than the radius, the radius is updated to that distance and the process is continued till only one point is left in the sphere. That will be the closest constellation point to the received point. If a point

is not found initially, then the sphere radius is incremented and the same process is followed [8].

III. ASTC-MIMO-OFDM SYSTEM

In this section we first describe the system model of ASTC-MIMO-OFDM chain, then the channel model is introduced, and finally the decoder structure is described.

A. System Model

We consider a coherent system over a frequency-selective correlated Rayleigh fading MIMO channel with two transmit and received antennas (Nt=Nr=2). The overall schematic diagram of ASTC-MIMO-OFDM transceiver is depicted in Fig. 2.

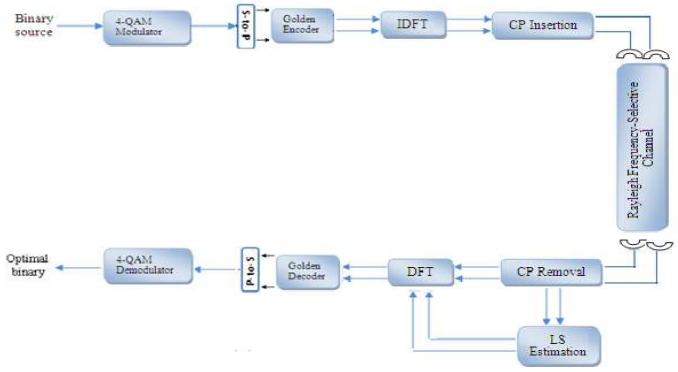


Fig. 2. Block diagram of Coded ASTC-MIMO-OFDM system

The same stages applied to information data in the first system model (section II) will be applied to this system.Before transmission over the two antennas, a conversion to a serial stream of the ASTC output is done. The Nt streams are then fed to Nt OFDM modulators, which uses an IFFT module, with Nfft subcarriers and a cycle prefix (CP) of length Nc. The Nt vectors of length Nfft+Nc are transmitted over a frequency and time selective MIMO channels. In order to avoid ISI, the CP length Nc is assumed to be longer than the largest multipath delay spread.

B. Channel model

We assume that the ASTC-MIMO-OFDM symbols are transmitted over a time and frequency selective Rayleigh channel and that the channel taps remain constant during a packet transmission. Consequently, the channel impulse response (CIR) between q^{th} transmitting antenna and p^{th} receiving antenna is modeled by a tapped delay line as

$$h_k^{p,q} = \sum_{l=0}^{L-1} h_k^{p,q}(l)\delta(k-l) \qquad (7)$$

where $h_k^{p,q}(l)$ is the l^{th} path from the q^{th} transmitting antenna to p^{th} receiving antenna at time k and L is the largest order among all impulse responses. The channel taps sequence $\{ h_k^{p,q}(l) \}$ is a correlated complex Gaussian process with zero mean and the same variance σ_h^2 and the autocorrelation function

$$E\left\{h_k^{p,q}(l)\left[h_{k-k'}^{m,n}(l')^*\right]\right\} = \rho_{Rx}^{(m,p)}\rho_{Tx}^{(n,q)} J_0(2\pi f_m k')\delta(l-l')$$

where J_0 is the Bessel function with zero order, f_m is the normalized Doppler shift, $\rho_{Rx}^{(m,p)}$, $\rho_{Tx}^{(n,q)}$ refers respectively to the correlation coefficient between the received antenna (m,p) and the transmitted antennas (n,q). To obtain a correlated Rayleigh fading channel, the autocorrelation function of $\{ h_k^{p,q}(l) \}$ process is given by

$$R_h = \sigma_h^2 \exp(j2\pi f_c k) J_0(2\pi f_m k) \qquad (8)$$

We can thus express the MIMO-OFDM received signal in a matrix notation as

$$y_k = h_k x_k + w_k \qquad (9)$$

Where x_k is k^{th} MIMO-OFDM symbol, w_k represents the AWGN at time k with $N_r N_{fft}$ i.i.d. elements and h_k is the equivalent channel matrix represented as

$$h_k = \begin{pmatrix} h_k(0) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ h_k(L-1) & \cdots & h_k(0) \end{pmatrix} \qquad (10)$$

where $h_k(l) = [h_k^{p,q}(l)]$ are the Nt*Nr matrices for l=0…L-1.

C. Decoder structure

At the receiver, after removing the CP, the signal is transformed back to the frequency domain by the mean of a DFT process. We can express the received frequency-domain signal as

$$Y = HX + W \qquad (11)$$

where W is the frequency domain noise with zero mean and variance σ_w^2, X is the frequency domain data matrix and H is a block diagonal matrix with Nt*Nfft frequency response of the channel matrix given by

$$H = \begin{pmatrix} H(0) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & H(Nt*Nfft-1) \end{pmatrix} \qquad (12)$$

Where n^{th} block H(n), n=0…Nt*Nfft-1, represents the MIMO-OFDM channel gain at the n^{th} subcarrier and can be written as

$$H(n) = \sum_{l=0}^{L} h_k(l) \exp(-j2\pi \frac{nk}{Nfft}) \quad (13)$$

We can express the restored useful data X as

$$\hat{X} = H^{\vdash} Y \quad (14)$$

where $(.)^{\vdash}$ denotes the pseudo-inverse operator.

As mentioned in equation 11 the restitution of signal needs the knowledge of the channel response which is generally unknown. In this section, we present a channel estimation method for OFDM systems using pilot symbols [9]. For MIMO-OFDM systems, pilots are inserted in both time and a frequency domain as it is shown in Fig. 3. Let us denote XP the vector of length P whose elements are the pilot symbols.

Based on the LS criterion, channel estimation method at pilot location, is given by

$$\hat{H}_p = (X_p)^{\vdash} Y_p \quad (15)$$

Then channel frequency response estimation at non-pilot positions can be done by interpolating the channel estimates at neighboring pilot symbol positions. Several efficient interpolation techniques for OFDM channel estimation have been investigated in [9]. In this work, we use the linear interpolation for its simplicity.

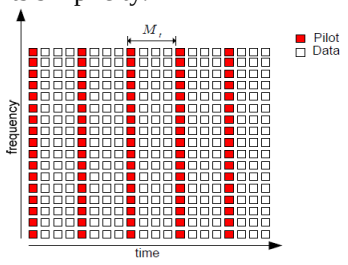


Fig. 3. Frequency and time domain insertion of pilot symbol

Once the channel effect is compensated, the decision variable $\hat{X}$ is passed for decoding. Zero Forcing sub-optimum decoder is used in this work, to reduce the numerical complexity without significant performance loss.

A serial to parallel module, at each DFT output, is used to reshape the signal $\hat{X}$. Then, we provide the sequences $\hat{X}_t$ and finally restitute the information sequence $\hat{S}$ given by

$$\hat{S} = \Phi_t^{-1} \hat{X}_t \quad (16)$$

IV. PERFORMANCE OF ASTC-MIMO-OFDM SYSTEM UNDER SPATIALLY CORRELATED MIMO CHANNELS

In practice, the MIMO channels between different antennas are often correlated which is called spatial correlation. As a result shown by information theory [11], the channel capacity can be substantially reduced for spatially correlated MIMO channels. It has been also shown that more the spatial correlation increases, more the BER decreases for the same SNR. So what is the impact of spatial correlation on the ASTC code design performances?

The capacity analysis of an ASTC-MIMO-OFDM system, under correlated Rayleigh frequency-selective channel, has been evaluated in [10]. It has been shown that the capacity increase almost linearly when the normalized Doppler frequency f_m decrease, and the match between the analytic and the simulated capacity, is more perfect for high Eb/No.

In this part we determine the analytic capacity expression of ASTC-MIMO-OFDM system under spatially correlated channel. To introduce the correlation effect in the system model, a correlated matrix should be generated according to [12]

$$vec(H_c) = C * vec(H) = R^{1/2} * vec(H) \quad (17)$$

$$R = CC^T$$

where H_c represent the correlated matrix of the MIMO channel. The correlation matrix of the MIMO channel, R, is obtained by the Kronecker product of the transmit and receive correlation matrices as follow [13]

$$R = R_{Tx} \otimes R_{Rx} \quad (18)$$

Using the properties of the Kronecker product, we have

$$H_c = R_{Tx}^{1/2} H (R_{Rx}^{1/2})^T \quad (19)$$

We can thus re-express the received frequency-domain signal in a matrix notation as

$$Y = H_c X + W \quad (20)$$

$$H_c = \begin{pmatrix} H_c(0) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & H_c(Nt * Nfft - 1) \end{pmatrix} \quad (21)$$

$$H_c(n) = \sum_{l=0}^{L} R_{Tx,l}^{1/2} h_k(l) R_{Rx;l}^{1/2} \exp(-j2\pi \frac{nk}{Nfft}) \quad (22)$$

First of all we should determine the mutual information between transmitted code word X and received vector Y. Then we can deduce the analytic expression of the information capacity of ASTC under spatially correlated MIMO channel for a given channel matrix H_c. The mutual information is given by

$$I(X,Y) = \mathrm{H}(Y) - \mathrm{H}(Y/X) = \mathrm{H}(Y) - \mathrm{H}(W) \quad (23)$$

Where $\mathrm{H}(Y)$ denotes the entropies of multivariate distribution Y. The capacity is obtained by maximizing the mutual information

$$C = \max(I(X,Y)) = \max(\mathrm{H}(Y)) - \mathrm{H}(W) \quad (24)$$

So we have to determine the exact expression of $\mathrm{H}(Y)$ to obtain the capacity expression. It has been shown in [10] that the expression of the capacity can be expressed as

$$C = \log_2 \det(I_{Nfft*Nt} + \frac{1}{\sigma_W^2} Q_X {H_c}^H H_c)\ bits/s/Hz \quad (25)$$

The covariance matrix Q_X can be expressed as [10]

$$Q_X = \{\Phi\Phi^H\} I_{Nfft*Nt} \quad (26)$$

where Φ is a copy of Φ_t (Nfft*Nt/4) times.

Then the capacity expression will be [10]

$$C = \log_2 \det(I_{Nfft*Nt} + \frac{1}{\sigma_w^2} \Phi\Phi^H {H_c}^H H_c) \quad (27)$$

If we derive the exact expression of $\Phi\Phi^H$ we get $I_{Nfft*Nt}$.

Exploiting the fact that H is a block diagonal matrix, we can write

$$H_c {H_c}^H = \begin{pmatrix} \|H_c(0)\|^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \|H_c(Nt*Nfft-1)\|^2 \end{pmatrix} \quad (28)$$

$$= \|H_c\|^2 = Nfft*Nt\|H_c(n)\|^2 \forall 0 \le n \le Nfft*Nt-1$$

Whereas

$$\|H_c(n)\|^2 = \left\| \sum_{l=0}^{L} R_{Tx,l}^{1/2} h_k(l) R_{Rx;l}^{1/2} \exp(-j2\pi \frac{nk}{Nfft}) \right\|^2$$

$$\le \sum_{l=0}^{L} \left\| R_{Tx,l}^{1/2} h_k(l) R_{Rx;l}^{1/2} \exp(-j2\pi \frac{nk}{Nfft}) \right\|^2$$

$$\le \sum_{l=0}^{L} \rho^2 \|h_k(l)\|^2 = \rho^2.L.P_h^2 \quad (29)$$

Where $\rho^2 = \sum_{l=0}^{L} \left\| R_{Tx,l}^{1/2} R_{Rx;l}^{1/2} \right\|^2$ and P_h denotes the power to each channel coefficient in time domain.

At this stage, the last capacity expression can be bounded by

$$C = \log_2 \det(I_{Nfft*Nt} \left[1 + \frac{1}{\sigma_w^2} \|H_c\|^2 \right])$$

$$= \log_2 (\left[1 + \frac{1}{\sigma_w^2} Nfft.Nt.\|H_c\|^2 \right]^{Nfft*Nt}) \quad (30)$$

$$\le \log_2 (\left[1 + \frac{1}{\sigma_w^2} Nfft.Nt.\rho^2.L.P_h^2 \right]^{Nfft*Nt})$$

$$\le Nfft.Nt.\log_2 (\left[1 + \frac{1}{\sigma_w^2} Nfft.Nt.\rho^2.L.P_h^2 \right])$$

By averaging over the capacity of Nfft*Nt narrowband channels we derive the following capacity

$$C \le \log_2 (\left[1 + \frac{1}{\sigma_w^2} Nfft.Nt.\rho^2.L.P_h^2 \right])\ bits/s/Hz \quad (31)$$

V. SIMULATION RESULTS

To investigate the performance of the proposed space time code, a series of Monte Carlo simulations were carried out. We first present coded ASTC system performances over correlated Rayleigh fading channel. Then we present the performance of the ASTC MIMO-OFDM system.

A. BER performance for coded ASTC system

Fig. 4 presents a comparison of Brute Force ML and Sphere decoding techniques. For Brute Force ML decoding technique, the BER is close to 10^{-1} for about 10 dB, however, in case of the Sphere decoding technique at 8 dB, the BER is about 10^{-1} so the gain is around 2dB for the Sphere decoding technique. The ML decoder suffers from a high complexity. These drawbacks can be addressed by the Sphere decoder technique, which helps reducing the BER.

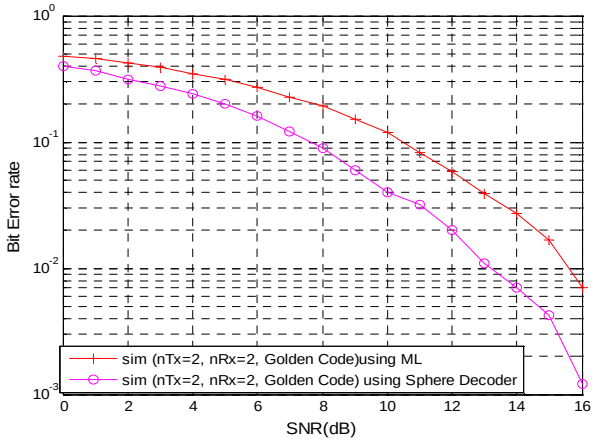


Fig. 4. Comparison of Brute Force ML and Sphere decoding techniques

Fig. 5 compares the BER performances of the Golden code and the classical Alamouti code. We can see that for a high SNR, the Golden code have a good performance in terms of BER. In the case of 2×1 transmission, the Golden BER remains close to Alamouti BER; for BER equal to 10^{-1} the gain is around 1dB for Golden code. By comparing this case to the 2×2 transmission case, the Golden code provides a gain of about 4dB for a BER equal to 10^{-1}. This gain comes from the fact that we are coding 4 symbols at the same time with Golden code, however we code only 2 symbols with Alamouti, then the gain still significant in terms of rate.

These results lead us to deal with real channel conditions mainly if we use a selective Rayleigh channel with unknown channel coefficients.

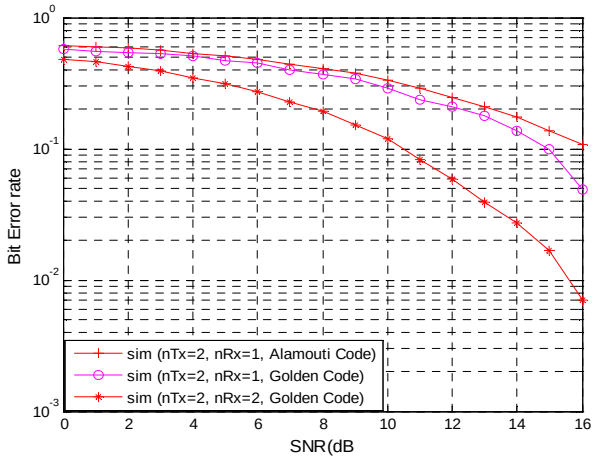


Fig. 5. Golden code versus Alamouti code

B. BER performance for ASTC-MIMO-OFDM system

In this sub-section we present a comparison of the BER performances between the ASTC code and the classical Alamouti code combined with MIMO-OFDM system.

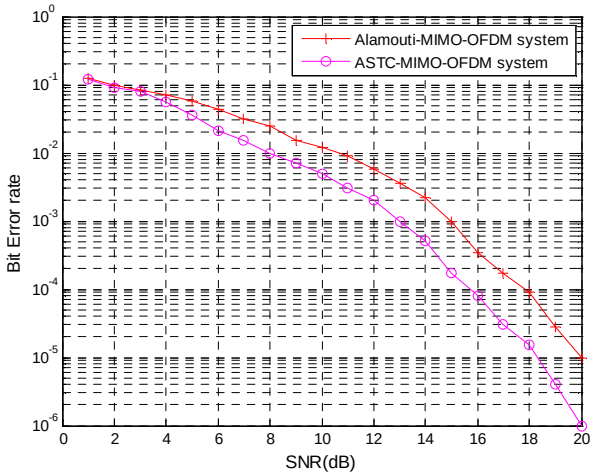


Fig. 6. ASTC-MIMO-OFDM versus Alamouti-MIMO-OFDM

Fig. 6 shows that the gain obtained by ASTC code is about 2 dB for a BER of 10^{-3}. The same raison for this gain: it comes from the fact that we are coding 4 symbols at the same time with ASTC code, however we code only 2 symbols with Alamouti.

C. Performance for ASTC-MIMO-OFDM system under spatially correlated MIMO channels

In this sub-section, we illustrate the results of the effect of spatial correlation of MIMO channel on the performances of ASTC-MIMO-OFDM architecture in terms of BER and capacity.

The BER Performances of ASTC code with respect to different spatial correlation is depicted in Fig. 7.

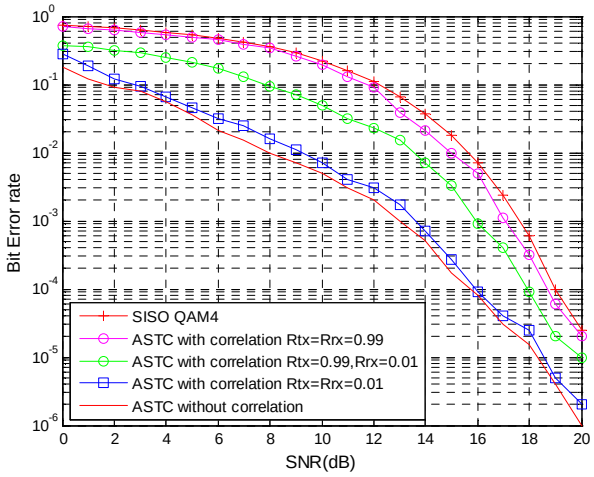


Fig. 7. BER Performances of ASTC code with respect to different spatial correlation

We can see that the correlation between sub channels degrade considerably the performances of ASTC-MIMO-OFDM system. In fact, for a fixed SNR of 10 dB for example, the BER of (R_{Tx}=R_{Rx}=0.99) is much important than the BER of (R_{Tx}=R_{Rx}=0.01). For a fixed value of correlation, the BER decrease when the value of the signal to noise ratio is

increased. We also notice that for (R_{Tx}=R_{Rx}=0.99), the BER is almost similar to uncoded SISO QAM-4 system. This can be explained by the fact that with strong correlation, the use of ASTC codes become less efficient and hence the importance of the spatial correlation on the hole performances.

The ergodic capacity as a function of SNR per receive antenna, for both the analytic capacity expression, derived in (30) and the simulated expression using (24), is shown in Fig. 8.

It can be observed that the strong correlation decrease considerably the ergodic capacity of the ASTC-MIMO-OFDM system compared to uncorrelated system. It can be observed also that the match between the analytic and the simulated capacity without or with low correlation, is more perfect for high SNR, mainly when SNR $\geqslant$ 16 dB.

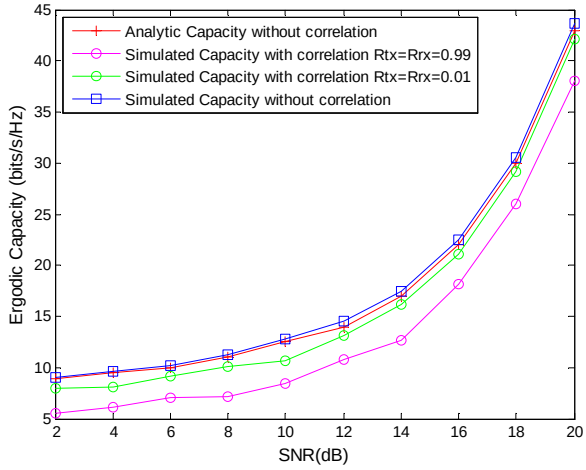


Fig. 8. Ergodic capacity of ASTC-MIMO-OFDM systen over spacially correlated and uncorrelated channels

VI. CONCLUSIONS

In this paper, we have proposed a MIMO transmission system, based on Algebraic space time coding which has good properties. Numerical results show that this code has a reasonable BER that outperforms the classical Alamouti code over correlated Rayleigh fading channel. In realistic multipath channel, frequency selectivity can be solved by the use of OFDM modulation. Numerical results show that ASTC codes maintain their properties and achieve good BER performances compared to the classical Alamouti MIMO-OFDM system. Finally, we have demonstrated that spatial correlation has an impact on ASTC code design: the strong spatial correlation decrease considerably the BER and capacity performances of the ASTC-MIMO-OFDM system.

REFERENCES

[1] S. Hara and R. Prasad, *Multicarrier Techniques for 4G Mobile Communications*, Artech House, June 2003.

[2] L. J. Cimini Jr., "Analysis andsimulation of a digital mobile channel using orthogonal frequency division multiplexing," IEEE Trans. Commun., vol. COM-33, pp. 665–675, July 1985.

[3] G. J. Foschini andM. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," Wireless Pers. Commun., vol. 6, no. 3, pp. 311–335, 1998.

[4] Y. Li, N. Seshadri, and S. Ariyavisitakul, "Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels," IEEE J. Select. Areas Commun., vol. 17, pp. 461–471, Mar. 1999.

[5] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," IEEE J. Select. Areas Commun., vol. 16, pp. 1451–1458, Oct. 1998.

[6] J. C. Belfiore, G. Rekaya, and E. Viterbo, "The golden code: a 2 x 2 fullrate space-time code with non-vanishing determinants," in Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on, Jun./Jul. 2004, pp. 310–310.

[7] M. O.Damen, H. E. Gama, G. Caire, "On Maximum-Likelihood Detection and the Search for the Closest Lattice Point",2003. *IEEE Trans. Inform. Theory*, vol. 49, no. 10, October 2003, pp.2389-2402.

[8] B. Hassibi and H. Vikalo, "On the Sphere-Decoding Algorithm I. Expected Complexity", AUGUST 2005. IEEE trans. on signal processing, vol. 53, no.. 8, pp.2806-2818.

[9] M. Morelli and U. Mengali, "A comparison of pilot-aided channel estimation methods for OFDMsystems," IEEE Trans. Signal Process., vol. 49, no. 12, pp. 3065–3073, Dec. 2001.

[10] A. Bannour, M. Ammari, S. Y, and R. Bouallegue, "On the capacity of ASTC-MIMO-OFDM system in a correlated Rayleigh frequency-selective channel," IEEE Veh. Technol. Conf., 2011.

[11] D. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multi-element antenna systems," in *Proc. Universal Pers. Commun. Conf.*, Oct. 1998.

[12] H. Ozcelik, M. Herden, W. Weichselberger, J. Wallace, and E. Bonek " Deficiencies of Kronecker MIMO radio channel model" *IEEE Electronics Letters.*, vol. 39, no. 16, pp. 1209–1210, August 2003.

[13] Kermoal, J. P., Schumacher, L., Pedersen, K. I., Mogensen, P. E., and Frederiksen, F "A stochastic MIMO radio channel model with experimental validation", Selected Areas in Communications, IEEE Journal on, vol 20, No 6, August 2002, pages 1211-1226.

Ines BEN HASSINE received the engineering degree and the M.Sc respectively in Telecommunications in 2008 and in Communications System in 2009 from the National Engineering School of Tunis (ENIT), Tunisia. From 2010 to 2013, she was a Research Associate with the Laboratory of InnovCOM (Innovation of COMunicant and COperative Mobiles), High School of Communication of Tunis, Tunisia. She is currently an Assistant at the High Institute of Computer Science Mahdia. His research interests include MIMO, OFDM systems, Space Time Code, UWB.

Ridha Bouallegue received the M.S degree in Telecommunications in 1990, the Ph.D. degree in Telecommunications in 1994, and the Habilitation a Diriger des Recherches (HDR) degree in Telecommunications in 2003, all from the National Engineer School of Tunis (ENIT), Tunisia. He is currently Professor in the National Engineer School of Tunis (ENIT) and Director of Research Laboratory Innov'COM / Sup'Com. His current research interests include mobile and satellite communications, Access technique, intelligent signal processing, CDMA, MIMO, OFDM and UWB system.

Distributed Mobility Control Schemes in the HIP-based Mobile Networks

Sang-Il Choi, Seok-Joo Koh

School of Computer Science and Engineering, Kyungpook National University, Daegu, Korea

overcycos@gmail.com, sjkoh@knu.ac.kr

***Abstract*— The Host Identity Protocol (HIP) has been proposed as an identifier-locator (ID-LOC) separation scheme, in which the 128-bit Host Identity Tag (HIT) is used as an ID and the IP address of the host is used as a LOC. In HIP, the mobility control operations are performed based on a centralized Rendezvous Server (RVS) that acts as a mobility anchor for mobile nodes, in which all the HIP control messages are passed through the RVS server. However, this centralized mobility scheme has some limitation, such as the service degradation by a point of failure and the overhead of centralized anchor. In this paper, we propose the two schemes for distributed mobility management (DMM): HIP-DMM-Push and HIP-DMM-Pull. From the numerical analysis, it is shown that the proposed DMM schemes can provide better performance than the existing centralized scheme, and that the pull-based distributed control scheme (HIP-DMM-Pull) provides the best performance among the candidate mobility schemes in terms of the processing overhead at the central RVS server and the HIP connection setup delays.**

***Keyword*—HIP, Rendezvous Server, Distributed management**

I. INTRODUCTION

WITH emergence of new types of wireless/mobile networks and wide popularity of smart phones, the number of mobile Internet users has been rapidly increasing. This mobile trend has caused a rapid growth of BGP routing table, as known as the routing scalability problem. [1].

To solve this problem the Host Identity Protocol (HIP) has been proposed in IETF [2], which splits the current IP address space into Host Identity (HI) and routing locator (IP address). An HI is encoded to Host Identity Tag (HIT) and both HI and HIT are not changed. If an end host wants to communication to another, it finds out other host's IPv4 address using the HIT of the other host. In HIP, the centralized Rendezvous Server (RVS) has the responsibility of this search process [3]. In the centralized scheme, all binding and first Initiate message (I1) are processed by a central RVS. However, the centralized scheme is vulnerable to several problems [4]. First, a single point of failure of central RVS may affect severe degradation of overall system performance and also the increased cost of network engineering. In addition, the centralized RVS has the risk of overhead by increasing of host that is stored in the RVS's HIT-IPv4 mapping table.

In this paper, we proposed the two distributed mobility management schemes in HIP-based mobile networks. The proposed schemes can be used to effectively provide the mobility support in HIP-based wireless/mobile HIP networks, compared to the existing centralized control schemes.

This paper is organized as follows. Section II describes the existing centralized schemes for HIP mobility control. In Section III, we propose the two distributed mobility control schemes. Section IV compares the existing and proposed schemes in terms of the HIP connection setup delays, and the processing overhead at the central RVS server. Section V concludes this paper.

Manuscript received February 26, 2013. This research was supported by Basic Science Research Program of NRF(2011-0026529) and ITRC support program of NIPA(NIPA-H0301-12-2004).

Sang Il Choi is with the School of Computer Science and Engineering, Kyungpook National University, Sangyeok 3-dong, Buk-gu, Daegu 702-701, Korea. e-mail: (overcycos@gmail.com).

Seok Joo Koh is with the School of Computer Science and Engineering, Kyungpook National University, Sangyeok 3-dong, Buk-gu, Daegu 702-701, Korea. e-mail: (sjkoh@knu.ac.kr).

II. EXISTING HIP SCHEME

In the existing scheme for HIP mobility control, all HIT-IPv4 mapping entries are stored at centralized RVS. Figure 1 shows the operation of existing HIP data transmission.

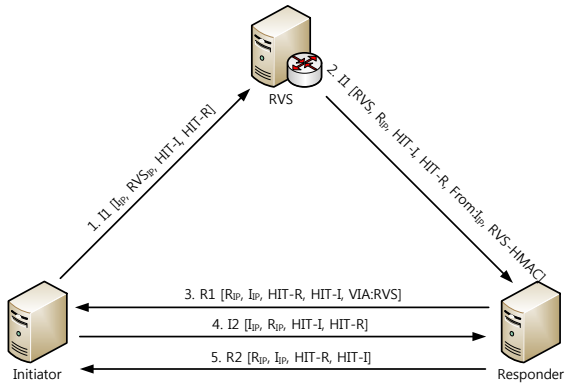


Fig. 1. Existing HIP data transmission

If Initiator wants to communicate with Responder, it sends the I1 message that contains the IP address of Initiator, IP address of RVS, HIT of Initiator, and HIT of Responder to centralized RVS, as described in Step 1. Next, RVS will search

the HIT-IPv4 mapping table. Then, RVS will forward I1 message that adds the RVS-HMAC for authentication to Responder (Step 2). On reception of I1 message, Responder responds with a R1 message that contains the IP address of Responder, IP address of Initiator, HIT of Initiator, and VIA:RVS for authentication (Step 3). After that, Initiator and Responder exchange the I2 message and R2 message (Step 4, 5). Now, communication between Initiator and Responder is possible, because Initiator knows the IP address of Responder.

In this paper, we will focus on only the inter-domain mobility control within a HIP-based mobile network, rather than the intra-domain mobility control.

Figure 2 describe the existing HIP data transmission schemes in the inter-domain case. First, Responder sends the Binding Update message to RVS for adding the HIT-IPv4 entry of Responder. Then, RVS responds with Binding ACK message to Responder. In this stage, if Initiator wants to communicate with Responder, it sends the I1 message to centralized RVS through Access Router (AR) and Gateway (GW) of Initiator. Then, centralized RVS will search the HIT-IPv4 table and forward to IP address of Responder through GW and AR of Responder. On reception of I1 message, Responder responds with R2 message to Initiator directly. After that, Initiator and Responder exchange the I2 message and R2 message through each AR and GW of Initiator and Responder.

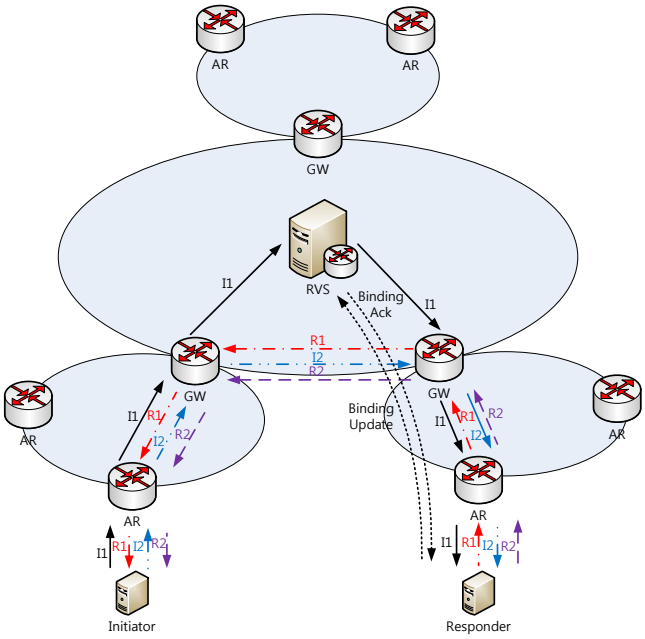


Fig. 2. Existing HIP schemes

III. Proposed DMM Schemes

A. Overview

In the proposed schemes, each GW has the Distributed Rendezvous Server (D-RVS) functionality, and stores the information on HIT and IPv4 for hosts.

Each of the proposed schemes updates the HIT-IPv4 table in the different way. In HIP-DMM-Push, if initiator is attached to a domain, the GW of Initiator sends the Binding Update message that contains the HIT and IP address of Initiator to the other D-RVS by multicast. On reception of this Binding Update message, each D-RVS updates their HIT-IPv4 table and then responds with Binding ACK message. This is called the 'push' operation, which is similar to the legacy routing protocol (e. g., OSPF) mechanism.

In HIP-DMM-Pull, on the other hand, each Binding Update message of Responder is completed by the D-RVS of Initiator, and only the D-RVS of Responder updates the binding entry. After that, if Initiator wants to communicate with Responder, the GW of Initiator sends the I1 message that is the first message of mapping query to the other GW through multicast for HIT-IPv4 mapping entry. Then, only the GW of Responder responds with R1 message that contains the IP address of Responder. From this R1 message, the D-RVS of Initiator can update the Responder's information in the HIT-IPv4 table.

The existing HIP scheme is a centralized scheme, in which all Binding Update and Mapping Query messages are processed by centralized RVS.

The proposed HIP-DMM-Push scheme is a distributed scheme, in which each GW performs the RVS functionality. In this scheme, the Binding Update with RVS is not performed. Instead, each D-RVS will send its Binding Update message to other D-RVS by multicast, when a new Responder is attached to the network. From this process, this scheme does not need to query for HIT-IPv4 mapping to other D-RVS, because the D-RVS already knows the associated information.

The proposed HIP-DMM-Pull scheme is also a distributed scheme with GW acting as RVS. The Binding Update is processed at D-RVS of the host. Then, each D-RVS sends an I1 message to other D-RVS to find the IP address of host.

B. HIP-Push

Figure 3 show the HIP-DMM-Push operations.

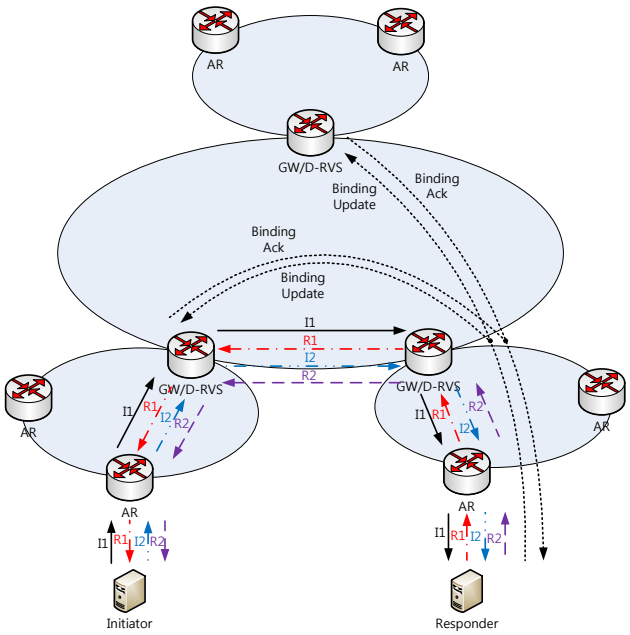


Fig. 3. HIP-DMM-Push operations

In the figure, when Responder is attached to a new domain, its HIT will be bound to its AR and GW/D-RVS. Then, GW of Responder will send (or push) the Binding Update message that contain the HIT and IP address of Responder to other GW/D-RVS by multicast. Every GW/D-RVS will update its HIT-IPv4 address, based on the Binding Update received from the GW/D-RVS of Responder. When Initiator wants to communicate with Responder, Initiator sends the I1 message to GW/D-RVS of Initiator. Then GW/D-RVS looks up the HIT-IPv4 table to find the IP address of Responder. If the IP address is found, GW/D-RVS forwards the I1 message to GW/D-RVS of Responder. Then, I1 message is forwarded to Responder. On reception of I1 message, Responder sends the R1 message to Initiator through GW/D-RVS of Responder and GW/D-RVS of Initiator. Next, Initiator sends the I2 message to Responder and receives the R2 message from Responder in the same way. Now, Initiator can send the data message to Responder directly.

C. HIP-Pull

Figure 4 shows the HIP-DMM-Pull operations.

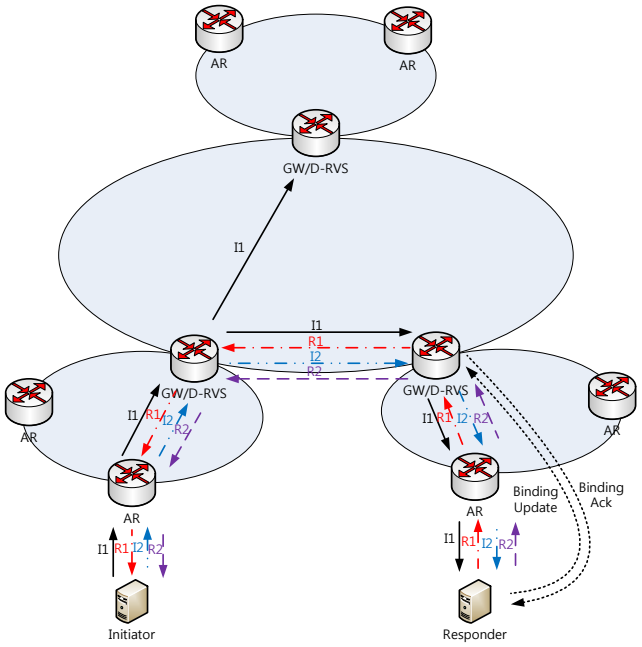


Fig. 4. HIP-DMM-Pull operations

In the figure, when Responder enters a domain, it is connected to AR and the Binding Update is processed at the GW/D-RVS of Responder. Now, the Initiator sends the I1 message to the GW/D-RVS of Initiator. Then, GW/D-RVS forwards the I1 message to other GW/D-RVS to find the IP address of Responder. Then, only the GW/D-RVS of Responder will forward the I1 message to Responder, the other GW/D-RVS will discard this message. On reception of I1 message, Responder sends the R1 message to Initiator through GW/D-RVS of Responder and GW/D-RVS of Initiator. Next, Initiator sends the I2 message to Responder and receives the R2 message from Responder in the same way. Now, Initiator can send the data message to Responder directly.

IV. Performance Analysis

To evaluate the performance of the proposed mobility schemes, we analyze the delay of the total transmission of first data and the overhead of GW/D-RVS. We compare the total transmission delay cost and the number of HIT-IPv4 table entry of GW/D-RVS for the existing scheme (HIP) and the proposed schemes (HIP-DMM-Push, HIP-DMM-Pull).

A. Analysis Model

Initiator and Responder are located within the different domain (i.e., Initiator is a mobile host and Responder is a static host), as illustrated below in the Figure 5.

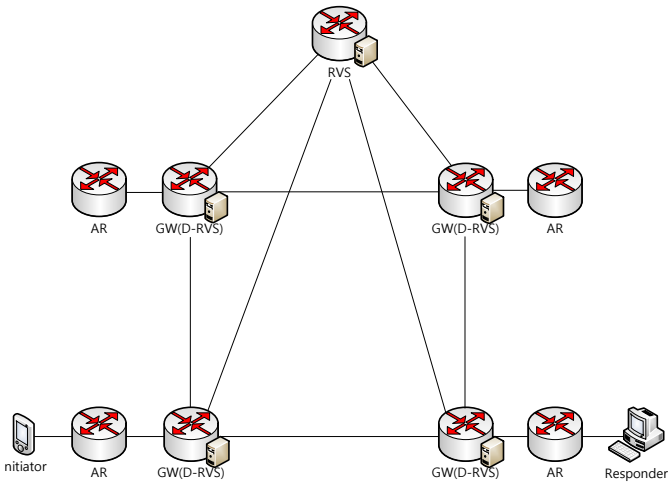


Fig. 5. Network model for numerical analysis

The binding update delay and the data delivery delay are denoted by BUD and DDD, respectively. Then the total delay (TD) is represented as TD=BUD+DDD. Then, for the processed signal at each D-RVS, we calculate the Signal Overhead (SO) that is sum of processed signal when Binding Overhead (BO) and Query Overhead (QO) process is performed.

B. Delay and Overhead Analysis

We define the parameters used for the analysis in Table 1.

TABLE I
PARAMETERS USED FOR DELAY AND OVERHEAD ANALYSIS

Parameter	Description
$\mathbf{T_{setup}}$	Node setup and HIT encoding delay
$\mathbf{S_b}$	Size of a 'b' packet
$\mathbf{N_{GW}}$	Number of GW in the Network
$\mathbf{H_{a-b}}$	Hop count between node a and b in the network
$\boldsymbol{\gamma}$ $\mathbf{(H_{GW\text{-}GW}/H_{GW\text{-}RVS})}$	Ratio of $H_{GW\text{-}GW}$ over $H_{GW\text{-}RVS}$
$\mathbf{P_c}$	Processing cost of node c for binding update or lookup

1) HIP

If the Responder enters a network, the HIT encoding process begins. We assume that this operation takes roughly T_{setup}. Then, Responder sends the Binding Update message to centralized RVS through GW and AR of Responder and receives the Binding Ack message. This operation takes $2S_{control} \times \{H_{AR\text{-}GW} + H_{GW\text{-}RVS} + (H_{host\text{-}AR}/2)\}$. On reception of Binding Update message, the centralized RVS updates their HIT-IPv4 table. This operation takes P_{RVS}. We assumed that the processing cost of RVS is proportional to the total number of active hosts in the domain ($N_{AR} \times N_{Host/AR}$) in the log scale by using a tree-based data structure to implement the database. Accordingly, the binding update delay of HIP can be represented as follows.

$$BUD_{HIP} = T_{setup} + 2S_{control} \times \{H_{AR\text{-}GW} + H_{GW\text{-}RVS} + (H_{host\text{-}AR} / 2)\} + \log_2(N_{host}/N_{RVS}) \quad (1)$$

In HIP, the data delivery delay for Initiator to Responder can be calculated as follows. First, Initiator sends the I1 message to centralized RVS through AR and GW of Initiator. Then, centralized RVS will look for the IP address of HIT of Responder in the HIT-IPv4 table, which takes $\log_2(N_{host}/N_{RVS})$. After that, RVS will forward the I1 message to Responder through AR and GW of Responder. This operation takes $2S_{I1} \times (H_{host\text{-}AR} + H_{AR\text{-}GW} + H_{GW\text{-}RVS})$. After that, Initiator and Responder exchange R1, I2, and R2 message through AR and GW of Initiator and Responder. Then, the first data packet will be forwarded from Initiator to Responder in the same way. This operation takes $(S_{R1} + S_{I2} + S_{R2} + S_{data}) \times (H_{host\text{-}AR} + H_{AR\text{-}GW} + H_{GW\text{-}GW} + H_{GW\text{-}AR} + H_{AR\text{-}host})$. Thus, the data delivery delay of HIP can be represented as follows.

$$DDD_{HIP} = 2S_{I1}(H_{host\text{-}AR} + H_{AR\text{-}GW} + H_{GW\text{-}RVS}) + (S_{R1} + S_{I2} + S_{R2} + S_{data}) \times (2H_{host\text{-}AR} + 2H_{AR\text{-}GW} + H_{GW\text{-}GW}) + \log_2(N_{host}/N_{RVS}) \quad (2)$$

So, we obtain the total delay of HIP as

$$TD_{HIP} = BUD_{HIP} + DDD_{HIP}$$

Now, to analyze the overhead of RVS, we will calculate the number of signals processed by RVS. For simplicity, we assume that the hosts are equally distributed in each domain. When hosts enter the network, they send a Binding Update message to the centralized RVS to update the HIT-IP address mapping entry. From this operation, the number of processed signal of RVS is N_{Host}. Then, if N_{Host} want to communicate with other hosts, they sends an I1 message to the other host through RVS. From this operation, the RVS shall process N_{Host}. From BO and QO, we can calculate the SO of HIP.

$$SO_{HIP} = BO_{HIP} + QO_{HIP} = 2N_{Host}$$

2) HIP-DMM-Push

If the Responder enters a network, HIT encoding process begins. We assume that this operation takes roughly T_{setup}. Then, Responder sends the Binding Update message to D-RVS of Responder and D-RVS will forward the Binding Update message to other D-RVS by multicast. This operation takes $2S_{control} \times \{H_{AR\text{-}GW} + H_{GW\text{-}GW} + (H_{host\text{-}AR}/2)\}$. On reception of Binding Update message, each D-RVS updates their HIT-IPv4 table. This operation takes $\log_2(N_{host})$, since each D-RVS has all host's mapping information from fowarded Binding Updated message. Accordingly, the binding update delay of HIP-DMM-Push can be represented as follows.

$$BUD_{HIP\text{-}DMM\text{-}Push} = T_{setup} + 2S_{control} \times \{H_{AR\text{-}GW} + (H_{host\text{-}AR}/2) + H_{GW\text{-}GW}\} + \log_2(N_{host}) \quad (3)$$

In the HIP-DMM-Push, the data delivery delay for Initiator to Responder can be calculated as follows. First, Initiator sends the I1 message to D-RVS of Initiator through AR and GW of Initiator. Then, D-RVS will look for the IP address of HIT of Responder in their HIT-IPv4 table, which takes $\log_2(N_{host})$. After that, D-RVS of Initiator will forward the I1 message to Responder through GW/D-RVS and AR of Responder. On reception of I1 message, Responder sends the R1 message to Initiator directly. Then Initiator and Responder exchange the I2 and R2 message in the same way. This operation takes $(S_{I1} + S_{R1} + S_{I2} + S_{R2} + S_{data}) \times (2H_{host\text{-}AR} + 2H_{AR\text{-}GW} + H_{GW\text{-}GW})$. Thus, the data delivery delay of HIP-DMM-Push can be represented as follows.

$$DDD_{HIP\text{-}DMM\text{-}Push} = (S_{I1} + S_{R1} + S_{I2} + S_{R2} + S_{data}) \times (2H_{host\text{-}AR} + 2H_{AR\text{-}GW} + H_{GW\text{-}GW}) + \log_2(N_{host}) \quad (4)$$

So we obtain the total delay of HIP-DMM-Push as

$$TD_{HIP\text{-}DMM\text{-}Push} = BUD_{HIP\text{-}DMM\text{-}Push} + DDD_{HIP\text{-}DMM\text{-}Push}$$

Now, we will calculate the number of signals processed by one D-RVS. When hosts enter the network. they send a Binding Update message to D-RVS that located at same domain to update the HIT-IP address mapping entry. In this case, the number of processed signal of one D-RVS is N_{Host}/N_{GW}. Then, each D-RVS forward the I1 message of host to other D-RVS by multicast. In this case, the number of processed signal of one D-RVS is $N_{Host}\text{-}N_{Host}/N_{GW}$. From these operations, BO of HIP-DMM-Push is N_{Host}. After that, if N_{Host} want to communicate with other hosts, they sends a I1 message to D-RVS of same domain. Since each D-RVS knows the location of other host, the I1 message is passed directly. From this operation, QO of HIP-DMM-Push is N_{Host}/N_{GW}. From BO and QO, we can calculate the SO of HIP-DMM-Push.

$$So_{HIP\text{-}DMM\text{-}Push} = BO_{HIP\text{-}DMM\text{-}Push} + QO_{HIP\text{-}DMM\text{-}Push} = N_{Host} + (N_{Host}/N_{D\text{-}RVS})$$

3) HIP-DMM-Pull

The Binding Update operations are performed as follows. When responder enters a network, HIT encoding process starts. We assume that this operation takes roughly T_{setup}. Then Responder sends the Binding Update message to D-RVS of Responder. This operation takes $2S_{control} \times \{H_{AR\text{-}GW} + (H_{host\text{-}AR}/2)\}$. At that time, the D-RVS of Responder updates their HIT-IPv4 mapping table. This operation takes $\log_2(N_{host}/N_{D\text{-}RVS})$, since each D-RVS stored only HIT-IPv4 mapping information of host that belongs to D-RVS's domain. Accordingly, the binding update delay of HIP-DMM-Pull can be represented as follows.

$$BUD_{HIP\text{-}DMM\text{-}Pull} = T_{setup} + 2S_{control} \times \{H_{AR\text{-}GW} + (H_{host\text{-}AR}/2)\} + \log_2(N_{host}/N_{D\text{-}RVS}) \quad (5)$$

In the HIP-DMM-Pull, the data delivery delay for Initiator to Responder can be calculated as follows. First, Initiator sends the I1 message to D-RVS of Initiator through AR and GW of Initiator. Then, D-RVS of Initiator forward the I1 message to other D-RVS by multicast. This operation takes $S_{I1}(H_{host\text{-}AR} + H_{AR\text{-}GW})$. After that, each D-RVS look for the IP address of HIT of Responder in their HIT-IPv4 table, which takes $\log_2(N_{host}/N_{D\text{-}RVS})$. If D-RVS find out that entry, then it will forward the I1 message to Responder through AR of Responder, and other D-RVSs discard the I1 message. This operation takes $S_{I1}(H_{host\text{-}AR} + H_{AR\text{-}GW} + H_{GW\text{-}GW})$. On reception of I1 message, Responder sends the R1 message to Initiator directly. Then, Initiator and Responder exchange the I2, R1, and I2 message, and next, Initiator sends the data packet to Responder directly. This operation takes $(S_{R1} + S_{I2} + S_{R2} + S_{data}) \times (2H_{host\text{-}AR} + 2H_{AR\text{-}GW} + H_{GW\text{-}GW})$. Thus, the data delivery delay of HIP-DMM-Pull can be reprented as follows.

$$DDD_{HIP\text{-}DMM\text{-}Pull} = S_{I1}(2H_{host\text{-}AR} + 2H_{AR\text{-}GW} + H_{GW\text{-}GW}) + (S_{R1} + S_{I2} + S_{R2} + S_{data}) \times (2H_{host\text{-}AR} + 2H_{AR\text{-}GW} + H_{GW\text{-}GW}) + \log_2(N_{host}/N_{D\text{-}RVS}) \quad (6)$$

So, we obtain

$$TD_{HIP\text{-}DMM\text{-}Pull} = BUD_{HIP\text{-}DMM\text{-}Pull} + DDD_{HIP\text{-}DMM\text{-}Pull}$$

Now, we will calculate the number of signal processed by one D-RVS. When hosts enter the network, they send a Binding Update message to D-RVS that located at same domain to update the HIT-IP address mapping entry. From this operation, the BO is $N_{Host}/N_{D\text{-}RVS}$. Then, if N_{Host} want to communicate with other hosts, they sends a I1 message to D-RVS of same domain. In this case, processed signal of each D-RVS is $N_{Host}/N_{D\text{-}RVS}$. After each D-RVS receive the I1 message, they forward this message to other D-RVS by multicast. This process is takes N_{Host}-$N_{Host}/N_{D\text{-}RVS}$. From this operation, the QO is N_{Host}. So we can calculate the SO of HIP-DMM-Pull.

$$SO_{HIP\text{-}DMM\text{-}Pull} = BO_{HIP\text{-}DMM\text{-}Pull} + QO_{HIP\text{-}DMM\text{-}Pull} = (N_{Host}/N_{D\text{-}RVS}) + N_{Host}$$

C. Numerical Results

Based on the cost analysis given in the previous section, we now compare the numerical results. For numerical analysis, we set the parameter values, as shown in Table 2, which are partly obtained from the results given in [5].

TABLE II
PARAMETER VALUES USED FOR DELAY AND OVERHEAD ANALYSIS

Parameter	Default	Minimum	Maximum
T_{setup}		100	
S_{I1}, S_{R1}, S_{I2}, S_{R2}		1	
$S_{control}$, S_{data}		1	
$N_{D\text{-}RVS}$	4	1	256
N_{host}	400	100	8000
$H_{host\text{-}AR}$		1	
$H_{AR\text{-}GW}$		1	
$H_{GW\text{-}GW}$		3	
$H_{GW\text{-}RVS}$		10	
γ	0.3	0.1	1.5

Figure 6 shows the impact of hop count ratio GW-GW over GW-RVS on total transmission delay. From the figure, we can see that the distributed mobility control schemes have lower total transmission delay until the ratio is equal to 0.8 and the HIP-DMM-Pull scheme shows the lower delay than HIP-DMM-Push. When the ratio is higher than 1.2, delay of distributed mobility control schemes are higher than existing HIP scheme. However, it is not a critical weakness. In general network, average distance between centralized server and gateway is farther than the distance between two gateways. So we are focus on the value that is lower than 1. In this general situation, we are sure that two distributed mobility control schemes can show the better performance than existing HIP scheme.

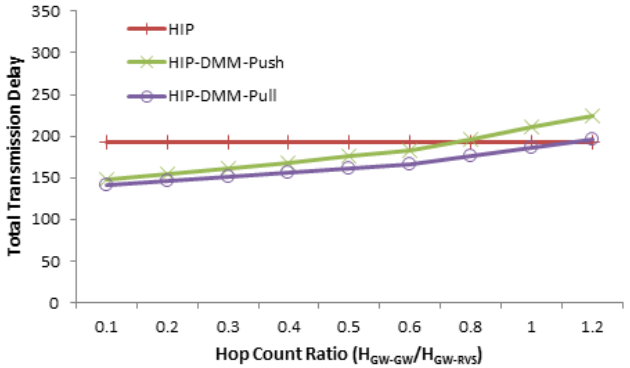


Fig. 6. Impcat of hop count ratio on total delay

Figure 7 and 8 show the impact of number of host and D-RVS on total transmission delay. From this figure, we can see that the two distributed mobility schemes show lower total transmission delay than existing HIP scheme. In HIP-DMM-Push, although each D-RVS has the same entry with RVS of existing HIP, there

is no query operation to get the Responder's IP address. So, this scheme can perform with low delay. In HIP-DMM-Pull, the I1 message's route is shorter than that of existing HIP. In addition, this scheme's HIT-IPv4 table has less entry than existing HIP scheme and HIP-DMM-Push. From these reasons, we can see that the two distributed mobility control schemes shows better performance than existing scheme and the HIP-DMM-Pull has the best performance.

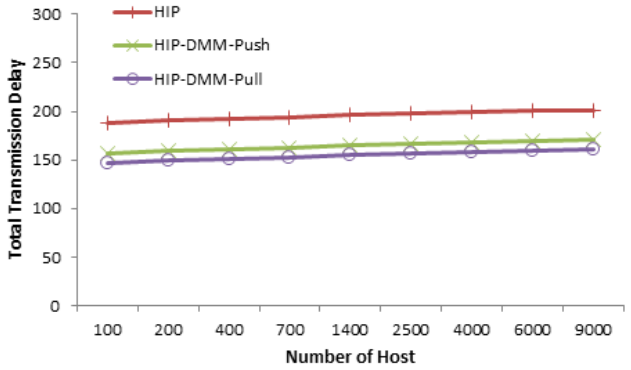


Fig. 7. Impact of number of host on total delay

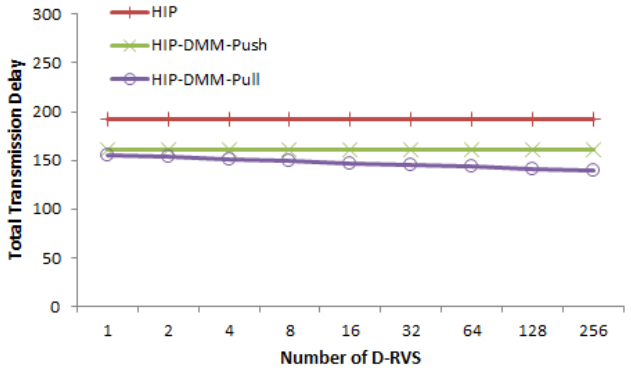


Fig. 8. Impact of number of D-RVS on total delay

Figure 9 and 10 show the impact of number of host and D-RVS. In HIP-DMM-Push, HIT-IPv4 table update of D-RVS is completed at binding update operation. So, the entry number of two schemes (HIP and HIP-DMM-Push) are the same. From this reason, the existing scheme and HIP-DMM-Push show the same performance. However, in HIP-DMM-Pull, each D-RVS has the MIP-IPv4 mapping entry of hos that is located in domain of GW. Accordingly, HIP-DMM-Pull has the best performance.

Figure 11 and 12 show the impact of number of host and D-RVS on processed signal in each D-RVS. From this figure, we can see that the proposed schemes shows better performance. In HIP, all Binding Update message and I1 message is processed at centralized RVS. However, in HIP-DMM-Push, only all Binding Update message is processed at D-RVS and each I1 message is processed at each D-RVS that located in same domain. In HIP-DMM-Pull, only all I1 message is processed at D-RVS and each Binding Update message is processed at each D-RVS that located in same domain. So, the two proposed schemes shows lower number of processed signal than existing HIP scheme.

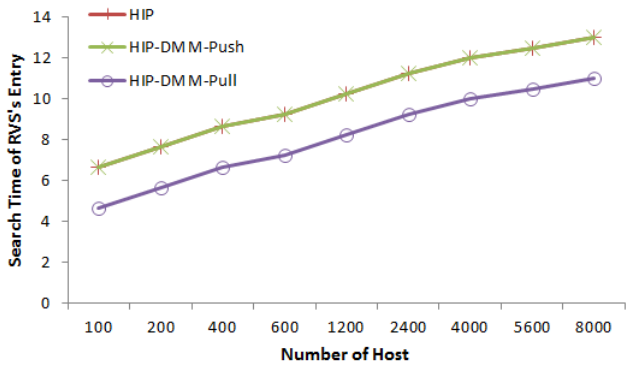


Fig. 9. Impact of number of host on search time of RVS

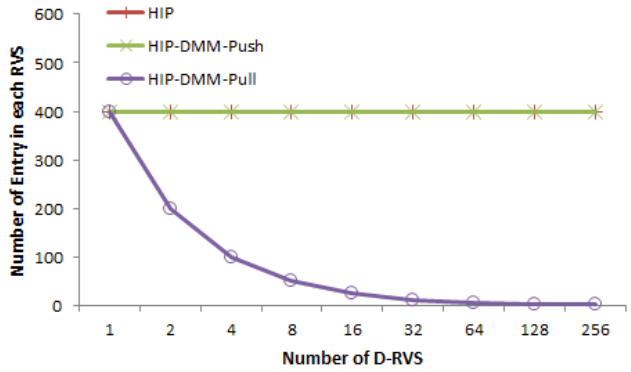


Fig. 10. Impact of number of D-RVS on performance

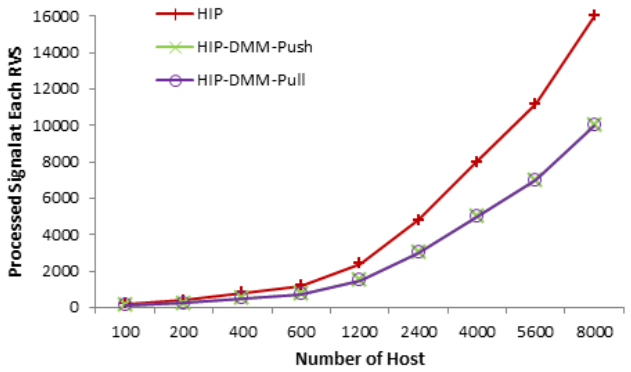


Fig. 11. Comparison of processed signal at each D-RVS

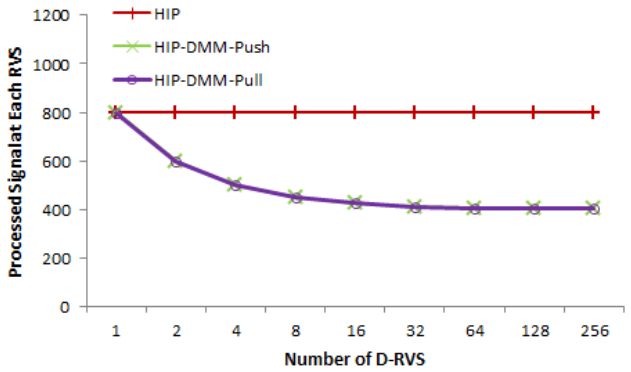


Fig. 12. Impact on the processed signal at each D-RVS

Figure 13 shows the impact of number of D-RVS in terms of generated signals in the network. In HIP, each host sends a Binding Update message and the I1 message to the centralized RVS. So, there is no affected by D-RVS. However, in HIP-DMM-Push, if host enters the network, it sends a Binding Update message to D-RVS that located in the same domain. On reception of Binding Update message, D-RVS forwards the message to other D-RVS and receives Binding ACK message from each D-RVS. From this operation, in Binding Update operation, the signals are generated by twice number of D-RVS. In HIP-DMM-Pull, if host want to communicate with Responder, it sends the I1 message to D-RVS that located in the same domain. Then, D-RVS forwards this message to other D-RVS. On reception of I1 message, D-RVS that is located in same domain with Responder responds with R1 message. From this operation, in Binding Query operation, the signals are generated by the number of D-RVS. Because of this, existing HIP scheme shows best performance.

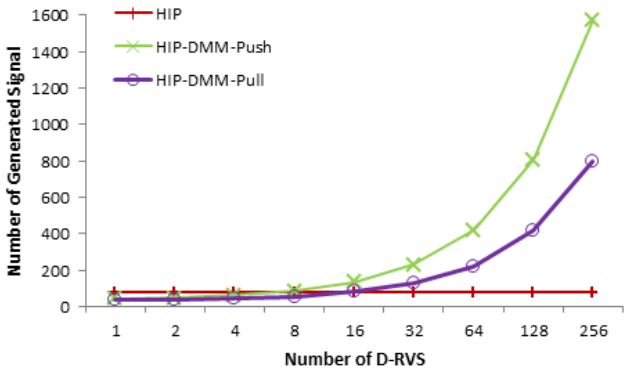


Fig. 13. Impact on the number of generated signal

V. Conclusions

In this paper, we propose the two schemes for distributed mobility management (DMM): HIP-DMM-Push and HIP-DMM-Pull. By numerical analysis, we compare the existing centralized RVS scheme and the two proposed distributed schemes, in terms of processing overhead at the central RVS server, and the HIP connection setup delays.

From the numerical results, on total transmission delay and processed signal at each RVS, we can see that the distributed mobility control scheme is better than the existing centralized RVS scheme. In particular, HIP-DMM-Pull scheme gives the best performance among all schemes. This is because HIP-DMM-Pull scheme's D-RVS has less HIT-IPv4 mapping entry than HIP-DMM-Push and Existing-HIP scheme.

References

[1] Morgan Stanley. Internet Trends. Available: http://www.morganstanley.com/institutional/techresearch/pdfs/Internet_Trends_041210.pdf, April 2010.

[2] R. Moskowitz, et al.., Host Identity Protocol (HIP), IETF RFC 5201, April 2008.

[3] J. Laganier, et al., Host Identity Protocol (HIP) Rendezvous Extension, IETF RFC 5204, April 2008.

[4] H. Chan, Problem statement for distributed and dynamic mobility management, IETF Internet Draft, draft-chan-distributed-mobility-ps-05.txt, October 2011.

[5] Y. H. Choi and T. M. Chung, "Using Correspondent Information for Routing Optimization Scheme on Proxy Mobile IPv6," Journal of Networks, Vol. 5, No. 8, pp. 984-989, August 2010.

Sang Il Choi received B.S and M.S. degrees in Engineering from Kyungpook National University in 2010 and 2012, respectively. Since March 2012, he enters to the Ph.D. program. His current research interests include mobility control and Future Internet such as Locator Identifier Separation Protocol (LISP) and Host Identity Protocol (HIP). Now, he studies the distributed mobility management in Future Internet.

Seok Joo Koh received B.S and M.S. degrees in Management Science from KAIST in 1992 and 1994, respectively. He also received Ph.D. degree in Industrial Engineering from KAIST in 1998. From August 1998 to February 2004, he worked for Protocol Engineering Center in ETRI. Since March 2004, he has been with the school of Electrical Engineering and Computer Science in the Kyungpook National University as an Associate Professor. He has published over 25 international journal papers with IEEE, Elsevier, and Springer-Verlag. His current research interests include mobility control in Future Internet, mobile SCTP, and mobile multicasting. He has also participated in the International Standardization as an editor in the ITU-T SG13 and ISO/IEC JTC1/SC6.

Comparative Study on Cooperative Communications in the Upper Layers of Ad Hoc Networks

Jaeshin Jang*, Sunghong Wie**

*Department of Information & Communications Engineering, INJE University, Korea

**Second Division of Information Technology, The Cyber University of Korea, Korea

icjoseph@inje.ac.kr, satwie@gmail.com

***Abstract*—In this study, we compare the cooperative MAC protocols, the routing protocols, and the transport protocols, all of which support cooperative communications. Numerical results and comparisons of three cooperative MAC protocols with reactive helper node selection mechanisms are presented. Research trends, as well as the advantages and disadvantages of related research, on cross-layer cooperative MAC protocols, cooperative routing, and transport protocols are described.**

***Index Terms*—Ad hoc, cooperative, cross-layer, helper node, MAC**

I. Introduction

THE goal of wireless communications is to attain a high transmission rate and end-to-end communication reliability by combating wireless channel impairments and a limited radio spectrum. By employing space diversity, cooperative communications were introduced in the physical layers of mobile ad hoc networks to overcome these limitations. The idea of space diversity being used in cooperative communications, also called cooperative diversity, is slightly different from the traditional space diversity (e.g., MIMO) because in cooperative communications single-antenna mobile nodes transmit cooperatively as a virtual antenna array. Figure 1 explains the concept of cooperative communications. As shown in Figure 1, there are three communication nodes in a sample network: the source, the destination, and the helper node. The source node sends data to the destination node; however, due to a bad wireless channel, the source node can only send data at 2 Mbps. If any communication node that is located between the source node and the destination node accepts a role as a helper node, then the source node and the helper node can send data at 11 Mbps. As a result of this helper node's contribution, system throughput has been increased. In addition, since the destination node receives data from both the source node and from the helper node, it gains space diversity [1].

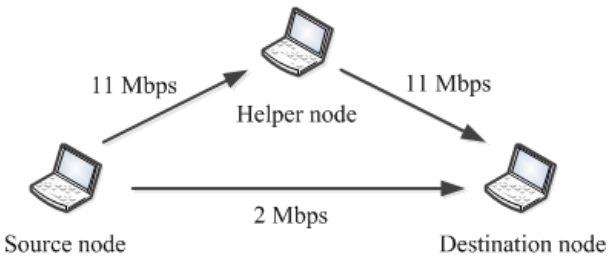


Fig. 1. Example of a cooperative communication.

Since the first studies on cooperative communications in physical layers, numerous studies have examined the physical layers' aspect. However, study results on MAC protocols and upper layers are scant, particularly when compared to research on physical layers. Therefore, a systematic study of these areas is required [12]. Instead of examining physical layers, this paper will conduct a trend survey of related research on cooperative communications in upper layers.

Section II is a survey of related studies, while Section III provides a comparative analysis of cooperative MAC protocols, routing protocols, and transport protocols that are used for cooperative communications. Section IV provides the conclusion.

II. Related Work

To date, the majority of cooperative MAC protocols that have been studied have been based on the IEEE 802.11 wireless LAN standard with various transmission rates. Determining how to decide which helper nodes to use in cooperative MAC protocols is one of the most important issues. According to when source nodes select helper nodes, the helper node selection mechanisms can be classified into two schemes: proactive and reactive schemes. The rDCF scheme [3] uses a proactive helper node selection scheme. Every mobile node

Manuscript received March 7, 2013. (Write the date on which you submitted your paper for review.) This work was supported in part by the 2012 NRF research grant (2012-R1A1A2041831).

J. Jang is with the Inje University, Gimhae, Gyeongnam, 621-749 KOREA (corresponding author to provide phone: 82-55-320-3520; fax: 82-55-322-6275; e-mail: icjoseph@inje.ac.kr).

S. Hong is with The Cyber University of Korea, Seoul, 110-800 KOREA (e-mail: satwie@gmail.com).

should maintain its relay table by overhearing control frames. The source node, which has data to send, first searches for one proper helper node from the relay table, and then transmits data to the destination node using two-hop communication through the helper node. The CoopMAC scheme [4] has many points of similarity to the rDCF scheme and maintains its CoopTable for selecting helper nodes in a proactive manner. However, these two schemes only use a merit of MAC protocol, but not cooperative diversity from the physical layer.

References [5]-[7] demonstrate cooperative MAC protocols with a reactive helper node selection scheme; in particular, references [5][6] are particularly cross-layer cooperative MAC protocols that receive support from the physical layer. The reactive scheme fulfils its role only after source and destination nodes exchange their RTS and CTS frames. In a cross-layer triple busy tone multiple access (CTBTMA) scheme [5], candidates for helper nodes calculate the utility function after overhearing the RTS and the CTS frame, at which point they send a busy tone. The utility function implies the effective transmission rate through the current wireless channel; thus, the bigger the utility function, the longer the busy tone. Therefore, the candidate node whose utility function is the largest is chosen as the final helper node and the node sends the ready to help (RTH) frame to both the source node and to the destination node. Reference [6] proposed a cross-layer cooperative MAC protocol where the candidate helper nodes calculate the composite cooperative transmission rate (CCTR) and then are classified into several groups according to the value of the CCTR. The contention procedure to select the best helper node is a two-step procedure: the first step takes place among different groups and the second step takes place among different members of one group. This competition procedure is known as the timer-based selection: the bigger the CCTR, the shorter the timer. Therefore, the candidate whose CCTR is the largest is defined as the helper node. After overhearing the RTS and the CTS frame, the cooperative relay-based auto rate (CRBAR) MAC protocol [7] uses a p-persistent back-off scheme to choose the helper node.

Numerous projects have researched cross-layer design. Reference [8] extended the CoopMAC to the cross-layer cooperative MAC protocol. The source node divides its channel slot in two and transmits only half the data in the first half. If the helper node receives the correct information, then it re-encodes and sends the rest of the data in the second half. Thus, the destination node gains space diversity by receiving half the coded data from the source node and the remaining half from the helper node. Reference [9] uses the CoopMAC as the fundamental access mode. According to the transmission rates between the source node, the helper node, and the destination node, there are four different transmission modes: basic access, RTS/CTS direct transmission, source-helper-destination transmission, and receiver maximal ratio combining. The receiver maximal ratio combining scheme is based on the cross-layer design between the physical layer and the MAC protocol. Reference [10] proposed the sender initiating concurrent cooperative MAC (SI-CCMAC) protocol for wireless LANs. This scheme is designed for the downlink of the AP's one-hop region and supports concurrent transmissions with the help of two-hop relaying. As is demonstrated in references [3][4], AP is required to maintain the helper table. The basic idea is that AP sends multiple data packets, which are used for different destinations, to multiple helper nodes; after this, the helpers simultaneously relay the data to the corresponding destinations.

In addition to studies on the cross-layer design of the physical layer and the MAC protocol, there have also been studies on cooperative routing and transport protocols. Reference [11] suggested a cooperative path access control (PAC) scheme that reserves the whole path, from the source node to the distant destination node, and then begins a pipeline communication. Reference [12] insisted that, in order to support various kinds of cooperation, network and transport layers are also needed to support cooperative communications.

Reference [13] proposed a cross-layer multilayer approach for supporting virtual multiple input single output (MISO) links. It includes a MAC layer any-cast protocol to find appropriate relay nodes, between which virtual MISO links are established. The primary path is constructed with the traditional ad hoc routing protocol, such as DSR or AODV, and relay nodes for establishing virtual MISO links are selected to shorten the primary path.

III. Analysis and Comparison

This section presents a comparative analysis of previous related works, which are described as three different topics. For each topic, the problems and benefits of each proposed scheme are explained in detail.

A. Helper node selection mechanism

There are two classifications of helper node selection mechanisms: a proactive mechanism and a reactive mechanism. The proactive mechanism is complex, causes lots of traffic in the network, and does not guarantee that the selected helper node, which is determined before the RTS transmission, is the best option for a time-varying wireless channel. In order to work properly, every mobile node should share the information in its helper table with its neighbours. However, no communication protocol has been suggested to make such sharing a possibility and any communication protocol that enables the possibility of this sharing may unnecessarily cause a heavy traffic load. Thus, the reactive mechanism is chosen as our study topic, and three cooperative MAC protocols using reactive mechanism are chosen for performance comparison.

The first protocol is the CTBTMA scheme [5]. It modifies the traditional dual busy tone multiple access (DBTMA) proposed for ad hoc networks into a cooperative MAC protocol by adding one additional busy signal. After exchanging the RTS and the CTS frame, the helper node candidates compute the proper channel transmission rate from the received SNR value in the channels between itself and the source or the destination node. If they find they can reduce the frame transmission time, then they participate in the selection of the helper node. Every participating candidate calculates the

following utility function U and sends a busy signal during a period that is proportional to its utility value.

$$U = W/(T_O + T_P) \quad (1)$$
$$T_P = W/r_{sh} + W/r_{hd} \quad (2)$$

After these helper node candidates finish sending the busy tone, they should monitor whether any other candidate continues sending a busy tone. If there is any continuing busy signal in the wireless channel, then these candidates are excluded from this helper node selection contest. In the end, the candidate whose utility value is the largest survives and this node sends the RTH frame to the source and the destination node. If the number of mobile nodes increases, then there will be more than one mobile node with the same utility value. Then there will be continual collisions in the helper node selection contest, which will consume both resources and time.

The cross-layer cooperative MAC protocol [6] also proposed a timer-based reactive helper node selection scheme. After overhearing the RTS/CTS frame transmitted by the source and the destination node, respectively, helper node candidates check whether they can contribute to increasing the effective transmission rate or not. When it is concluded that they can be helpful, they send a helper indication (HI) signal. They calculate CCTR R_h using equation (3), and they are then divided into several groups according to the value of the CCTR.

$$R_h = \frac{W}{W/R_{C_1} + W/R_{C_2}} = \frac{R_{C_1} R_{C_2}}{R_{C_1} + R_{C_2}} \quad (3)$$

This scheme uses a two-step contention mechanism. The first contention is inter-group and the second contention is intra-group or among the members within that group. This contention is carried out with mobile nodes' timers whose values are inversely proportional to their CCTR. Each helper node candidate triggers its timer after the CCTR calculation is finished. As soon as its timer expires, the candidate checks whether there is any signal, such as a group indication (GI) or a member indication (MI) signal. If any signal is found, then the candidate is excluded from this helper node selection contest. Otherwise, it sends a GI or an MI signal. If this transmission is successful, the candidate sends the RTH frame. In order to avoid a possible collision with other RTH frames, the related candidates retransmit their RTH frames in a mini-slot chosen randomly among succeeding K mini-slots.

The CRBAR scheme proposed in Reference [7] uses a reactive helper node selection mechanism in order to decide the proper helper node. First, after overhearing the RTS/CTS frame, all helper node candidates calculate the received SNR. After this, they begin the helper node selection process in each of the following time slots, using the p-persistent back-off scheme, where the probability p is decided using the following equation:

$$p = \frac{\delta}{1/r_{sh} + 1/r_{hd}} \quad (4)$$

In the above equation, δ is the system parameter and is closely related to system performance. However, a *p*-persistent back-off scheme does not guarantee that the selected helper node will be the best mobile node.

Reference [9] proposed a reactive helper node selection scheme, which is similar to the CTBTMA. The difference is that helper node candidates, which are determined to have smaller expected transmission times than those in the direct transmission, participate in the helper node selection procedure with an RTH frame rather than a busy signal. Therefore, this mechanism has the same problem, i.e. continuous collisions, when there is more than one candidate with the same expected transmission time.

Three different cooperative MAC protocols, suggested in references [5]-[7], are compared using a computer simulation. The simulation code is implemented in C++ language with the SMPL simulation tool [15]. For this performance evaluation, the IEEE 802.11b wireless LAN with an ad hoc mode is used as a sample network, and the system parameters used in the performance evaluation are described in Table 1. All mobile nodes are assumed to be randomly distributed in an 100m × 100m square communication area, and every mobile node is assumed to move independently in a random waypoint model [14]. It is also assumed that every source node always has packets to transmit in its buffer, which is called the saturated traffic model. This saturated traffic model is commonly used to derive the maximum system performance. In this paper, it is assumed that there is no frame transmission error for all frames, including the control and data frames. As performance measures, system throughput and channel access delay are used; system throughput is defined as the number of bits transmitted successfully during the whole simulation time and channel access delay is defined as the time period from the instant when the source node sends the CRTS frame to the instant when the source node receives the ACK frame successfully. The relation between the rates and the ranges is shown in Table 2 [5].

TABLE I
SYSTEM PARAMETERS

Parameters	Values	Parameters	Values
RTS	160 bits	SIFS	10 μs
CTS	112 bits	DIFS	50 μs
RTH	164 bits	CWmin	32 slots
ACK	112 bits	CWmax	1024 slots
DATA	1024 bytes	Basic rate	1 Mbps
Slot time	20 μs	MAC header	28 bytes
GI (MI)	3 (3)	K	4
δ	0.06	CRBAR mini-slot	6
Simulation time		1500 sec	
Transmission rate		1, 2, 5.5, 11 Mbps	

TABLE 2
TRANSMISSION RATES AND RANGES

Data rate	11 Mbps	5.5 Mbps	2 Mbps	1 Mbps
Range ($BER \leq 10^{-5}$)	48.2 m	67.1 m	74.7 m	100 m

Figure 2 and Figure 3 show the system throughput and the channel access delay of the CTBTMA scheme, respectively. It is shown in Figure 2 that the CTBTMA scheme has the best

performance when the number of source nodes is five. This is because five source nodes with the saturated traffic model are enough to make this communication area completely saturated, thus when there are more than five source nodes the probability of transmission collisions increases, which results a the decrease of system throughput. Figure 2 also shows that the CTBTMA scheme has the best system throughput when the number of helper nodes is two. Since the number of supportable transmission rates is limited (ex. 1, 2, 5.5, and 11 Mbps here), utility function defined in equation (1) for helper node candidates has a limited number of values. Therefore, as the number of helper nodes increases, the probability that there is more than one helper node candidate with the same value of utility function increases. This causes repeated collisions of helper node candidates in a helper node selection contest, which results in reduced system throughput.

It is shown from Figure 3 that channel access delay is closely related to the number of source nodes but not to the number of helper nodes. This means that the contention among source nodes with CRTS frames to seize a wireless channel is a more dominant factor to channel access delay than that among helper node candidates with RTH frames.

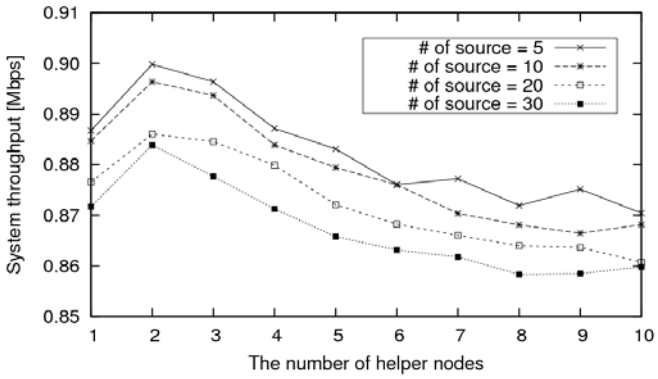


Fig. 2. System throughput of CTBTMA

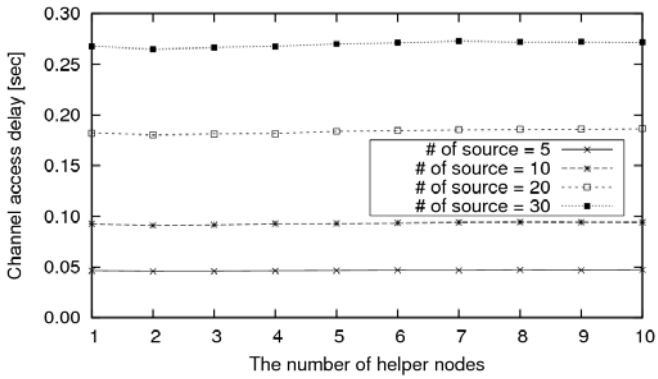


Fig. 3. Channel access delay of CTBTMA

Figure 4 and Figure 5 show the system throughput and the channel access delay of the cross-layer cooperative MAC protocol (for the sake of simplicity, called the cross-layer MAC from here on) suggested in Reference [6], respectively. It is shown in Figure 4 that this cross-layer MAC protocol has the best system throughput performance when the number of helper nodes is between 5 and 20, which is greater than the value 2 in the CTBTMA scheme. In addition, system throughput performance is greater than that of the CTBTMA scheme over its whole gamut. It is shown in Figure 5 that the channel access delay of the cross-layer MAC performance is slightly lower, but it has a pattern similar to the CTBTMA scheme.

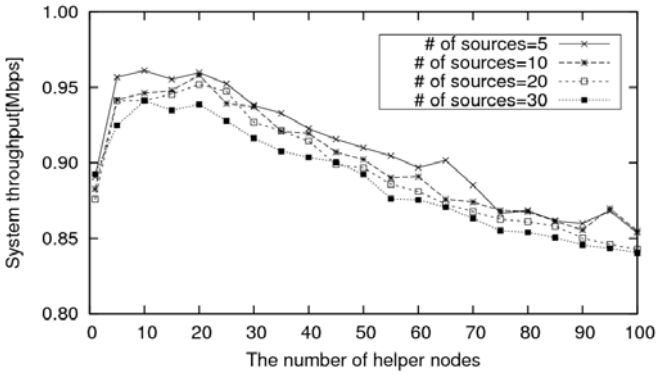


Fig. 4. System throughput of the cross-layer MAC [6]

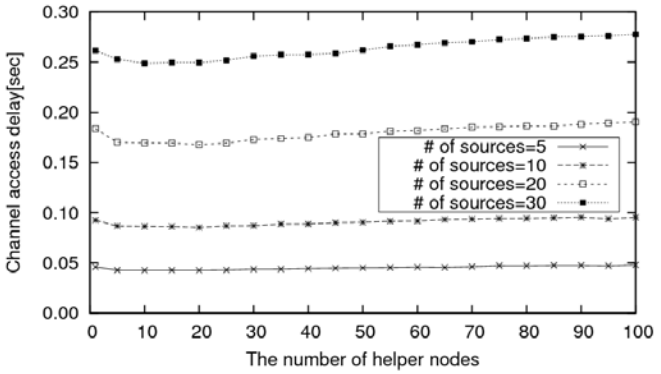


Fig. 5. Channel access delay of the cross-layer MAC [6]

Figures 6 ~ 8 show the system throughput and the channel access delay of the CRBAR MAC protocol suggested in Reference [7], respectively. The system parameter δ in equation (4) used to derive numerical results in Figure 6 and Figure 7 is 0.02. It is shown in Figure 6 that the system throughput of the CRBAR scheme tends to decrease rapidly compared to Figure 2 and Figure 4. It is also shown in Figure 7 that the channel access delay of the CRBAR scheme tends to increase a little more rapidly than the CTBTMA scheme and the cross-layer MAC scheme. Since the CRBAR scheme makes use of the p-persistent back-off scheme to select an appropriate helper node, its performance is greatly dependent on the system parameter δ in equation (4). The relationship between the system throughput and the system parameter δ is shown in Figure 8 when the number of helper nodes is 5 and 20, respectively. In this figure, no cooperation means the traditional IEEE 802.11 wireless LAN protocol without cooperation. It is shown in Figure 8 that the CRBAR has the best system throughput when the system parameter $\delta = 0.06$.

Therefore, the system parameter with $\delta = 0.06$ for the CRBAR scheme will be used for performance comparison of these three cooperative MAC protocols.

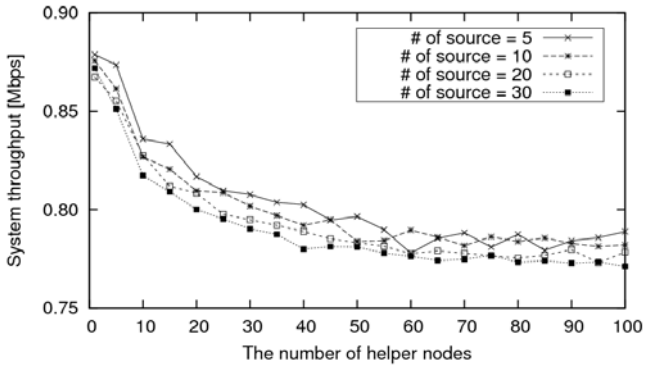


Fig. 6. System throughput of CRBAR ($\delta = 0.02$)

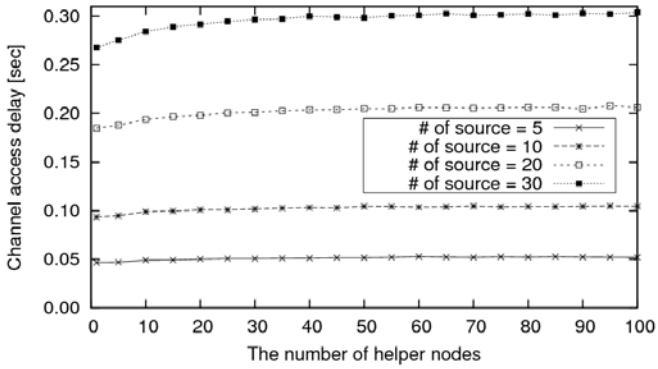


Fig. 7. Channel access delay of CRBAR ($\delta = 0.02$)

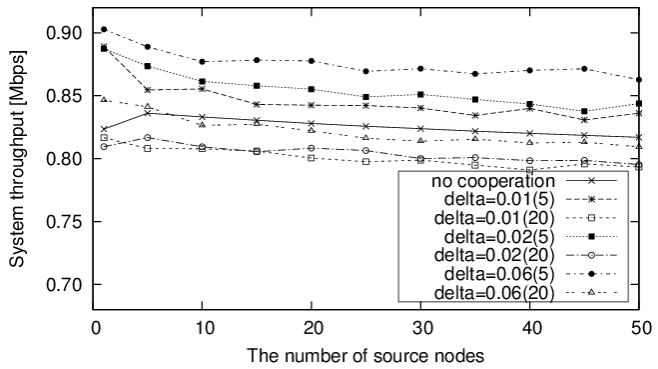


Fig. 8. System throughput of CRBAR for various δ

Figure 9 and Figure 10 show system throughputs for three cooperative MAC protocols in a different environment. First, system throughputs in Figure 9 are derived when there are five source nodes and δ=0.06, as is the case with the CRBAR scheme. This figure shows that the cooperative MAC protocol suggested in Reference [6] has the highest system performance, while the next highest is the CTBTMA scheme [5], and the third highest is the CRBAR scheme [7]. As the number of helper nodes increases, the success probability in choosing the correct helper node decreases because the number of helper node candidates with the same value in equation (1) or equation (3) increases.

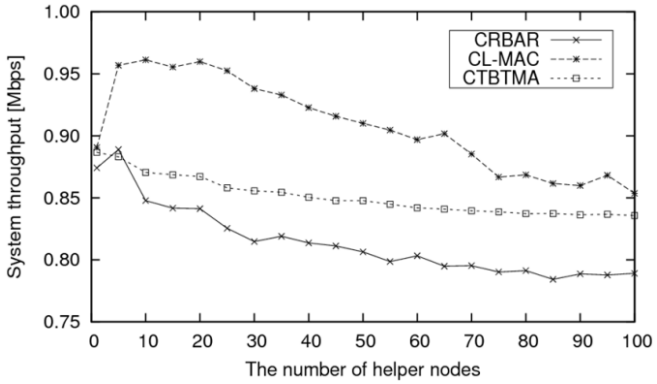


Fig. 9. System throughput of three schemes as a function of the number of helper nodes

Figure 10 shows the system throughput performance for three cooperative MAC protocols and the IEEE 802.11b MAC protocol without cooperation when the number of helper nodes is 10. It is shown that all three cooperative MAC protocols have better system performance than the MAC protocol without cooperation, and the throughput gap between Reference [6] and the others is large. Since, with the saturated traffic model, source nodes always have data frames in their buffers, this figure shows that the total system throughputs for the CTBTMA and the MAC protocol without cooperation reach the maximum value when there are approximately five source nodes.

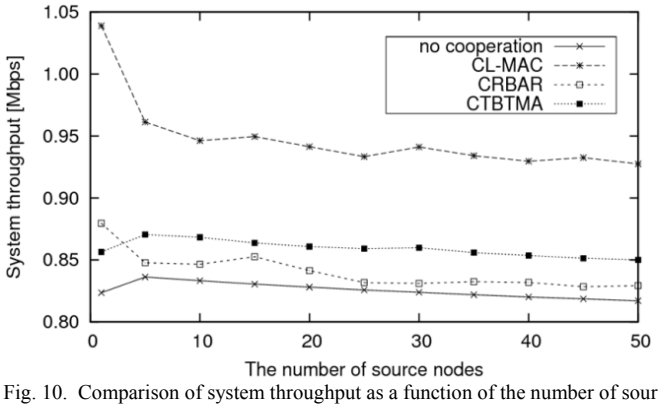


Fig. 10. Comparison of system throughput as a function of the number of sour ce nodes

B. Cross-layer design

Since the approach using only the physical layer or the MAC protocol has a limit for maximizing a merit of cooperative communications, the cross-layer design between two layers is required. There have been many studies on the cross-layer design, especially with a physical layer and the MAC protocol. Reference [8] extended the CoopMAC [4] to the cross-layer cooperative MAC protocol in order to enhance system performance. Each channel block, or packet, contains data (B) and parity bits (r), leading to the equation $N = B + r$ coded bits. The source node divides its channel block in two and transmits only half of its coded bits in the first half time slot. These bits

are received both by the destination node and the helper node. If the helper node has the correct information, then it re-encodes and sends the remaining half in the second half time slot. Thus, the destination node receives half the coded bits from the source node and the remaining bits from the helper node, creating space diversity. But there are two questions that result from using this approach. The first question is related to the fact that the first half of coded bits, which the helper node receives from the source node, may not contain enough parity bits because it is assumed that $r \geq B$, as is shown in Reference [8]. Thus, the first question is how can the helper node decide whether the received bits are erroneous or not without parity bits in the extreme case? The second question is that separating the assigned transmission period in two time slots is opposite to the design goal of the MAC protocol, the goal of which is to reduce transmission time by raising channel transmission rate.

Reference [9] also extended the CoopMAC scheme to the cross-layer MAC protocol. By using three transmission rates, one for the direct path and two for the two-hop path, this scheme uses four different transmission modes. This approach seems to be as good as the cross-layer design between the physical layer and the MAC protocol; however, mobile nodes should share lots of information with their neighbours. The SI-CCMAC protocol proposed in reference [10] for wireless LANs is designed for the downlink of the AP's one-hop region and supports concurrent transmissions with the help of two-hop relaying. However, it is questionable whether this concurrent transmission in an AP's region is possible, especially when the helper nodes are located near one another.

C. Cooperative approach in upper layers

Compared with current studies about the physical layers and the MAC protocols, studies on cooperative routing protocols or transport protocols are relatively scant. Reference [11] investigated the overheads and challenges that emerge from implementing cooperative communications in a realistic network. Cooperative PAC, which is proposed in Reference [11], is a CSMA-based scheme where the source node sends an RTS-like control frame and then the destination node responds with a CTS-like control frame. This exchange reserves the channel along the entire route for a specified period of time. For route selection in the cooperative PAC, a modified DSR routing protocol, which needs a cross-layer design between the routing and the MAC protocol, is proposed.

Reference [12] insisted that, in order to maximize the merit of cooperative communications, systematic studies on the network and transport layers are required. Compared to traditional one-link based communications, cooperative communications generate a new concept for links, which is a multi-path link between two neighbouring mobile nodes, and this is called a cooperative link. Therefore, cooperative routing protocols handling a sequence of cooperative links should be studied systematically. In this reference, to increase the achieved throughput with cooperation, aggregating the offered resources from different cooperating entities is considered. This kind of cooperative communication can be achieved through cooperative strategies in the network and transport layers. In this scenario, source and destination nodes use several IP addresses. As a result, a new transport layer approach for merging several IP addresses into one network address at each end point should be required, and the typical solution may be a transport layer protocol with a multi-homing feature.

IV. CONCLUSION

In order to maximize the benefits from cooperative communications, systematic research into whole layers is required. At the beginning of this research, a survey on cooperative communications in the upper layers, rather than physical layer, has been described. It includes the helper node selection mechanisms, the cross-layer cooperative MAC protocols, the network layer, and the transport layer supporting the cooperation. Their problems and solutions are described in order to decide future research directions. The performances of three cooperative MAC protocols with reactive helper node selection mechanisms are evaluated and compared to each other using a computer simulation.

REFERENCES

[1] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 74-89, October 2004.

[2] IEEE Std. 802.11b-1999, *Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications*, September 1999.

[3] H. Zhu and G. Cao, "rDCF: A relay-enabled medium access control protocol for wireless Ad Hoc Networks," *IEEE Trans. on Mobile Computing*, vol. 5, no. 9, pp. 1201-1214, September 2006.

[4] P. Liu, Z. Tao, S. Narayanan, T. Korakis, and S. S. Panwar, "CoopMAC: A cooperative MAC for wireless LANs," *IEEE J. of Selected Areas on Commun.*, vol. 25, no. 2, pp. 340-353, February 2007.

[5] H. Shan, et. al., "Cross-layer cooperative triple busy tone multiple access for wireless networks," *in Proc. of IEEE GLOBECOM-2008*, pp. 1-5, New Orleans, December 2008.

[6] H. Shan, "Cross-layer cooperative MAC protocol in distributed wireless networks," *IEEE Trans. on Wireless Commun.*, vol. 10, no. 8, pp. 2603-2615, August 2011.

[7] T. Guo, R. Carrasco, "CRBAR: Cooperative relay-based auto-rate MAC for multi-rate wireless networks," *IEEE Trans. on Wireless Commun.*, vol. 8, no. 12, pp. 5938-47, December 2009.

[8] F. Liu, T. Korakis, Z. Tao, and S. Panwar," A MAC-PHY cross-layer protocol for wireless ad hoc networks," *in Proc. of the IEEE Wireless Communications and Networking Conference(WCNC-208)*, pp. 1792-1797, Las Vegas, April 2008.

[9] Chunguang Shi et al., "CAC-MAC: A cross- layer adaptive cooperative MAC for wireless Ad hoc networks," *Int. Journal of Distributed Sensor Networks*, pp.1-9, 2012.

[10] Zhenqing Hu, Chen-Khong Tham, "SI-CCMAC: Sender initiated concurrent cooperative MAC for wireless LANs," *in Proc. of Int. Conference on Modeling and Optimization in Mobile, Ad hoc, and Wireless Networks(WiOPT-2009)*, 2009

[11] Justin Yackoski, et al., "Networking with cooperative communica- tions: Holistic Design and realistic evaluation," *IEEE Communications Mag.*, pp. 113~119, August 2009.

[12] W. Zhuang and M. Ismail, "Cooperation in wireless communication networks," *IEEE Wireless Communications*, pp. 10~20, April 2012.

[13] G. Jakllari, S. V. Krishnamurthy, M. Faloutsos, P. V. Krishnamurthy, and O. Ercetin, "A cross-layer framework for exploiting virtual MISO links in mobile ad hoc networks," *IEEE Trans. On Mobile Computing*, vol. 6, no. 6, pp. 579-594, June 2007.

[14] David B. Johnson and David A. Maltz, "Dynamic source routing in ad hoc wireless networks," *in Mobile Computing*, Chapter 5, pp. 153~181. Kluwer Academic Publishers, 1996.
[15] M. H. MacDougall, Simulating computer systems: Techniques and tools, The MIT Press, 1992.

Jaeshin Jang (M'90) received a B.S. degree in Electrical Engineering from Dong-a University, Korea, in 1990, and M.S. and Ph.D. degrees in Electric and Electrical Engineering from KAIST, Korea, in 1992 and 1998, respectively. From July 1997 to February 2002, he worked for Samsung Electronics Company. From August 2008 to July 2009, he was a visiting scholar at Iowa State University. Since March 2002, he has worked for INJE University, Korea. His major interests are wireless QoS, MAC, routing, and cooperative communications in wireless communications networks, including mobile WiMAX, ad-hoc networks, and mesh networks.

Sunghong Wie received B.S., M.S. and Ph.D. degrees in electrical engineering from KAIST, South Korea, in 1995, 1997, and 2001, respectively. From 2001 to 2008, he was with Samsung Electronics as a senior engineer. Currently, he is with the Division of Information Technology at The Cyber University of Korea. His recent research interests include mobility management and the future Internet.

Design of Small-Area Transimpedance Optical Receiver Module for Optical Interconnects

Jamshid Sangirov, Ikechi Augustine Ukaegbu, Nga T. H. Nguyen, Tae-Woo Lee, Mu-Hee Cho, and Hyo-Hoon Park

Photonic Computer Systems Laboratory, Electrical Engineering Department, KAIST, Daejeon, Korea 305-701

jamshid@kaist.ac.kr, aus20@kaist.ac.kr, nganth@kaist.ac.kr, twlee@ee.kaist.ac.kr, chophy@kaist.ac.kr, pakrhh@ee.kaist.ac.kr

***Abstract*—The development and miniaturization of electronic devices and components is pushing the system devices and their interconnecting interfaces to become even smaller. Thus, reducing the size of receiver (Rx) and transmitter (Tx) chips plays an plays an important role in designing a small-size optical modules utilized in o/e and e/o converters. Therefore, designing a small-area optical Rx may require intuitive solutions, such as building single-ended Rx and utilizing some of the advantages of differential Rx. Optical Rx should convert optical input signal to voltage output signal and provide sufficient gain and frequency operation for feeding to subsequent blocks including clock and data recovery circuit (CDR) and/or Serializer and Deserializer (SerDes). Therefore, we have designed a small-area transimpedance optical receiver (TIORx) using regulated-cascode (RGC) as an input stage which converts input photocurrent to voltage signal. The RGC block is connected to post amplifying stages to increase the overall transimpedance gain of the TIORx. The post amplifying gain stages utilizes two intersecting active feedback in order to increase the frequency operation in addition to increasing the gain of the proposed TIORx chip. The TIORx module is designed in a 0.13µm CMOS technology and works up to 10 Gbps data rate. The TIORx chip core occupies an area of 0.051mm^2 with power consumption of 16.9 mW at 1.3 V. A measured 3-dB bandwidth of 6.9 GHz was obtained for the TIORx module with a transimpedance gain of 60 dBΩ.**

***Keyword*—Optical receiver, small-area circuit design, bandwidth improvement, optical interconnections**

I. INTRODUCTION

As the data rate of internet traffic is increasing the electrical interconnection cannot provide reliable data at higher frequencies due to its limitations at high frequency operation [1]. Thus, optical interconnection is dominating the market because of its advantages such as low electromagnetic interference (EMI), negligible insertion loss with inner-channel crosstalk for multi-channel optical link and light weight comparing to electrical interconnection at high data rates. Hence, optical interconnection system is being widely utilized in high data rate video and data transmission systems [2]. However, optical interconnection also requires electrical-to-optical (e/o) and optical-to-electrical (o/e) converters. Designing a small-size o/e and e/o converters is important. Thus, reducing the size of receiver (Rx) and transmitter (Tx) chips plays an plays an important role in designing a small-size optical modules utilized in o/e and e/o converters. Since, the development and miniaturization of CMOS technology is pushing the design limits to design even smaller electronic devices and their interconnecting interfaces. Thereby, requiring the reduction of optical modules of o/e and e/or converters utilized in applications such as small form-factor pluggable (SFP) and system in package (SiP) systems for board-to-bard, chip-to-chip and rack-to-rack optical interconnection [3-4].

Designing of a compact optical Rx circuits for high-speed operation is important to reduce the signal propagation loss and reduce the active are used for chip design. Hence, designing of small-area optical Rx modules may require intuitive solutions, such as utilizing some of the advantages of differential Rx in designing of a single-ended Rx modules. Optical Rx chip should also be able to provide a sufficient gain and frequency operation with reduced size for feeding the signal to post-processing blocks such as CDR and/or SerDes [5]. In our previous work, we have analyzed the performance of single-ended and differential Rx in terms noise, power, size, and gain [6]. From comparison of single-ended and differential topology in table 1, the single-ended Rx chip size showed much smaller size. Hence, the size of Rx module can be significantly reduced by using a single-ended Rx. However, the data rate of single-ended Rx module was relatively small comparing differential topology. Therefore, we have tried to improve the performance single-ended Rx by increasing the data rate of Rx using he Rx chip [7].

In an optical Rx design, the transimpedance amplifier (TIA) plays the role of a front-end amplifier for amplifying the weak

Manuscript received March 12, 2013. This work was supported by the IT R&D program of MKE/KEIT [10039230, Development of bidirectional 40 Gbps optical link module with low power in Green Data Center for Smart Working Environment] and it was also supported by the Center for Integrated Smart Sensors funded by the Ministry of Education, Science and Technology as Global Frontier Project (CISS-2012366054191).

J. Sangirov, I. A. Ukaegbu, N. T. H. Nguyen, T.-W. Lee, M.-H. Cho, and H.-H. Park are with Electrical Engineering Department of Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, 305-701 (phone: +82-42-350-5453; e-mail: jamshid@kaist.ac.kr).

current signals generated from the photodiode (PD) and converting to voltage signal which would be fed to subsequent blocks (limiting amplifier or clock and data recovery circuit). The conventional Rx module consists of a TIA preamplifier block, limiting amplifier, and an output buffer. TIA design using inductive peaking for bandwidth enhancement and high data rate operation has been proposed in [8]. However, the excessive size of the inductor makes the total chip size big. An inductorless TIA has been designed in [9], where several shunt feedback TIAs connected in parallel were suggested for bandwidth improvement and chip size reduction. However, this design method has high power consumption due to the several TIAs deployed. In conventional Rx circuits, the TIA may not have enough gain and bandwidth to drive other circuits and as a result, may require the use of LA, equalizer [10-11], and other performance enhancement circuitry. However, due to the need for high data-rate and high performance operation, the need for the additional circuitry (LA, equalizer, etc.) will result to the increase in circuit complexity and thus, lead to increase in total chip area and increase in overall power consumption.

In this work, we propose a small area, high data-rate and inductorless TIORx which combines some of the advantages of both the TIA and Rx. While acting as a front-end amplifier for the weak signals from the PD, it also incorporates the gain and bandwidth enhancement advantages of an LA, in a conventional Rx. Thus, due to the absence of a dedicated LA stage, the TIORx module reduces the total chip area, while improving the gain and bandwidth performance. In our proposed design, the TIORx is made up of a regulated cascode (RGC) input stage; an inter-stage buffer; a gain stage with intercepting active feedbacks; and an output stage (which acts as a buffer). The RGC input stage has been utilized to reduce the input capacitance effect from the PD while improving the bandwidth [12]. The inter-stage buffer used between RGC and gain stage reduces parasitic capacitance at the output of RGC block, thereby improving bandwidth performance. The post amplifying stages utilizes active feedback only, making the chip less sensitive to process changes compared to a passive feedback system. The fabricated TIORx achieves a transimpedance gain of 60 dBΩ and occupies an area of 50.6 μm^2, which is quite small when compared to other high data-rate optical receivers of similar CMOS technology.

II. The TIORx Circuit Design

The TIORx is designed in a 0.13μm CMOS technology and the schematic of the TIORx chip is shown in Fig. 1. The input stage, which is the RGC, consists of transistors M_1 and M_2.The RGC input stage reduces the input impedance by the amount of its own voltage gain, which prevents the input pole from dominating the TIORx bandwidth and reduces capacitive effect of the PD [5]. Thus, the RGC circuit can be used effectively for CMOS integration as a front-end amplifier. The input impedance of RGC stage is given as [12]:

$$Z_{IN-RGC} = \frac{1}{gm_2(1+gm_1R_1)}, \quad (1)$$

where $1+gm_1R_1$ is the gain of the local feedback and with the product of the common gate stage, it behaves as a large transconductor $G_m=gm_2(1+gm_1R_1)$. Thus, the size of local feedback decides the amount of reduction of input parasitic capacitance effect for bandwidth determination. The RGC peaking frequency in frequency response is given as:

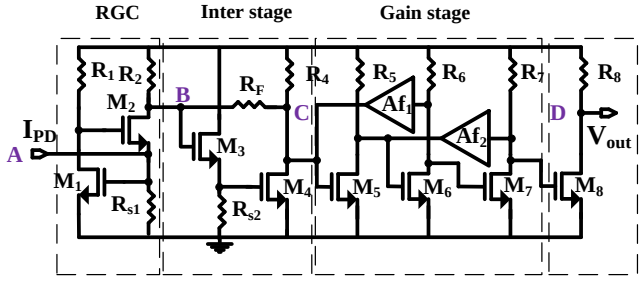


Fig. 1. (a) Schematic of TIORx.

$$f_{peak} = \frac{1}{2\pi R_1(C_{gs2}+C_{gd1})}. \quad (2)$$

At low frequency, the open-loop transimpedance gain of the TIORx is given as:

$$Z_{IN-GAIN} = -(R_2 \| R_F)\frac{gm_3R_{s2}}{(1+gm_3R_{s2})}\left(gm_4 \| gmf_1\right) \quad (3)$$
$$R_4\left(gm_5 \| gmf_2\right)R_5gm_6R_6gm_7R_7gm_8R_8$$

The TIORx is made up of four stages, namely, an input stage (which consists of the RGC); an inter-stage; a gain stage with intercepting active feedbacks; and an output stage (which acts as the buffer). The RGC block is important in the TIORx circuit as it affects the input noise and the stability of the whole TIORx circuit while delivering the input photo current to the output with increased gain. Thus, the design parameters have to be carefully chosen not to interfere with the input impedance for high frequency operation.

M_3 and M_4 make up the inter-stage stage of the TIORx, where high frequency operation should be maintained for delivering the converted input current to output voltage for the gain stage. The impedance at the drain of M_2 (of RGC stage) and M_4 (of inter-stage) are reduced by factor of *1+A* by R_F shunt feedback and correspondingly, the poles are sped up by the factor, *1+A*. A common-drain (CD) is placed at the drain of M_2 because the capacitive effect on bandwidth is small [12]. The inter-stage isolates the RGC input stage from the gain stage and also adjusts the input dc level from the RGC stage.

To increase the overall gain of the TIORx to higher output voltage levels, several common source (CS) amplifying stages (M_5 – M_7) have been utilized. However, placing these CS gain stages may reduce the bandwidth. Thus, to increase the bandwidth while maintaining a high gain level, two intersecting active feedback stages have been utilized [13]. The negative

active feedback utilized in the gain stage is different from the conventional resistive feedback which avoids the direct resistive load to the preceding transimpedance stage. Moreover, active devices suffer less process variations than passive devices during fabrication. Hence, adding the active feedback stages, Af_1 and Af_2, compensate by peaking at high frequencies. The high-frequency peaking occurs at Af_1 and Af_2 active feedback. The peaking of the first and second active feedbacks is given as:

$$f_{peak,f1} = \frac{1}{2\pi R_4 (C_{gd6} + C_{gdf1})} \tag{4}$$

$$f_{peak,f2} = \frac{1}{2\pi R_5 (C_{gd7} + C_{gdf2})} \tag{5}$$

As the circuit is more complex, we separate the circuit from point A to B, B to C and C to D as shown in Figure 1. Firstly, the poit from I_{in} to I_B is the core RGC block, front-end of TIA circuit, which has an important effect in the input noise and stability of the whole Rx circuit and delivers the input photo current to the output with increased gain. Thus, the design parameters have to be carefully chosen to interfere with the input impedance for high frequency operation. The transfer function can be derived as follows:

$$\frac{I_B}{I_{in}}(s) = \frac{1+\frac{sC_1}{gm_1}}{\left[1+\frac{s(C_{gs1}+C_{sb1})}{(1+gm_1R_1)gm_2}\right]\left[1+sR_1\left(C_1+C_{gs2}+C_{gd2}\right)\right]} \tag{6}$$

The next point is from I_B to V_C is inter-stage block, where the high frequency operation should be maintained for delivering the converted input current to output voltage for the gain stage. The impedance of drain M_2 and M_4 are reduced by factor of (1+A) by R_F shunt feedback and correspondingly the poles has been speed up by the factor (1+A). The common-drain (CD) is placed at the drain of M_2 because the capacitive effect on bandwidth is small [12]. The transfer function from B to C is with simplification of R_1 and R_f parallel resistors, is given as follows:

$$R_F^{'} = (R_1 \,||\, R_F)\frac{1}{1+\left(\frac{gm_3R_{s2}}{1+gm_3R_{s2}}\right)gm_4R_4}$$

$$\frac{V_C}{I_B}(s) = \frac{gm_3gm_4R_{s2}R_4R_F^{'}}{1+gm_3R_{s2}}\left(1+\frac{sCgs3}{gm3}\right) \tag{7}$$

$$\frac{1}{\left[1+\frac{s(R_F^{'}C_{gs3}+gm_3R_{s2}R_F^{'}C_1^{*}}{1+gm_3R_{s2}}\right]}\frac{1}{\left[1+\frac{s\left(C_{gs3}+C_{gs4}\right)}{gm_3}\right]}$$

The intersecting active feedback from V_C to V_{out} is the gain stage, where the voltage input has been increased to higher output voltage levels. Adding the negative active feedback increases the 3-dB bandwidth. The active feedback effects of Af_1 and Af_2 has been included in the transfer function and the equation is given as:

$$\frac{V_{out}}{V_C} = \frac{G_5(s)G_6(s)G_7(s)}{1+G_5(s)G_6(s)Gf_1(s)+G_6(s)G_7(s)Gf_2(s)} = \frac{G^3(s)}{1+2G^2(s)Gf(s)} \tag{8}$$

with the assumption $G_5(s) = G_6(s) = G_7(s) = \frac{GmR}{1+sRC}$, and $Gf_1(s) = Gf_2(s) = Gf(s) = \frac{GmfR}{1+sRC}$.

The last stage capacitance effect is loaded by input capacitance of output buffer. After combining the equations (6) to (8) we can write the transfer function of Rx as:

$$\frac{V_{out}}{I_{in}} = \frac{I_B}{I_{in}} * \frac{V_C}{I_B} * \frac{V_{out}}{V_C} =$$

$$\frac{1+\frac{sC_1}{gm_1}}{\left[1+\frac{s(C_{gs1}+C_{sb1})}{(1+gm_1R_1)gm_2}\right]\left[1+sR_1\left(C_1+C_{gs2}+C_{gd2}\right)\right]} \tag{9}$$

$$\frac{gm_3gm_4R_{s2}R_4R_F^{'}}{1+gm_3R_{s2}}\left(1+\frac{sCgs3}{gm3}\right)\frac{1}{\left[1+\frac{s(R_F^{'}C_{gs3}+gm_3R_{s2}R_F^{'}C_1^{*}}{1+gm_3R_{s2}}\right]}$$

$$\frac{1}{\left[1+\frac{s\left(C_{gs3}+C_{gs4}\right)}{gm_3}\right]}\frac{G^3(s)}{1+2G^2(s)Gf(s)}$$

The transimpedance gain of the TIORx can be obtained from equation (9) and is written as follows:

$$Z_T(0) = \frac{gm_3gm_4R_{s2}R_4}{1+gm_3R_{s2}} * \left[R_1 \,||\, \frac{R_F}{1+\left(\frac{gm_3R_{s2}}{1+gm_3R_{s2}}\right)gm_4R_4}\right] \tag{10}$$

The 3-dB bandwidth of the TIORx is affected by dominant poles at amplifying stages of gm_3, gm_5 and gm_6. Thus, the dominant poles can be described by the frequency response of the gain stages with transconductance of the dominant poles given by the time constant equations of (11) to (15):

$$\tau_{eq} = \tau_3 + \tau_5 + \tau_6 \tag{11}$$

$$\tau_3 = \left(R_F \,||\, \frac{1}{gm_3}\right)\underbrace{\left[C_{s3}+C_{gs4}+\left(1+gm_4R_4\right)C_{gd4}\right]}_{\alpha_1} = \left(R_F \,||\, \frac{1}{gm_3}\right)\alpha_1 \tag{12}$$

$$\tau_5 = \left(\left(R_4 + \frac{1}{gm_4}\right) \| \frac{1}{gmf_1}\right) \underbrace{\left[C_{s5} + C_{gs6} + (1 + gm_6 R_6) C_{gd6}\right]}_{\alpha_2} \quad (13)$$
$$= \left(\left(R_4 + \frac{1}{gm_4}\right) \| \frac{1}{gmf_1}\right) \alpha_2$$

$$\tau_6 = \left(\left(R_5 + \frac{1}{gm_5}\right) \| \frac{1}{gmf_2}\right) \underbrace{\left[C_{s6} + C_{gs7} + (1 + gm_7 R_7) C_{gd7}\right]}_{\alpha_3} \quad (14)$$
$$= \left(\left(R_5 + \frac{1}{gm_5}\right) \| \frac{1}{gmf_2}\right) \alpha_3$$

$$\tau_{eq} = \left(R_F \| \frac{1}{gm_3}\right)\left(\left(R_4 + \frac{1}{gm_4}\right) \| \frac{1}{gmf_1}\right) \quad (15)$$
$$\left(\left(R_5 + \frac{1}{gm_5}\right) \| \frac{1}{gmf_2}\right) \alpha_1 \alpha_2 \alpha_3$$

The 3-dB response of the TIORx can be obtained with τ_{eq} from $f_{3dB}=1/(2\pi\tau_{eq})$. Writing in terms of three dominant poles of TIORx, the three major poles would be $P_1=\tau_1$; $P_2=\tau_5$; $P_3=\tau_6$.

The input-referred noise of TIORx is described as follows:

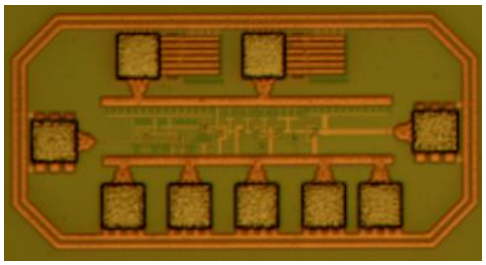
The input referred noise of Rx can be described as follows:

$$I_{n,eq} \cong I_{s1}^2 + I_{Rf}^2 + I_{R2}^2 + \left(\frac{s\left(C_{db2} + C_{gs3}\right)}{gm_3}\right)^2 \left(I_{d3}^2 + I_{R3}^2\right)$$
$$+ \left[\frac{R_1\left(1 + s\left(C_{in} + C_{gs1} + C_{sb2}\right)\right)}{\left(1 + gm_2 R_1\right) Rs}\right]^2 I_1^2$$
$$+ \left(\frac{s\left(C_{gs2} + C_{gd1}\right)}{gm_1}\right)^2 \left(I_{d2}^2 + I_{Rf}^2 + I_{R2}^2\right)$$
$$\cong 4kT\left(\frac{1}{R_{s1}} + \frac{1}{R_f} + \frac{1}{R_2}\right) + \frac{4kT\omega^2\left(C_{db2} + C_{gs3}\right)^2}{gm_3^2}$$
$$\left(\Gamma + \frac{1}{R_3}\right) + \frac{4kT\omega^2\left(C_{gs2} + C_{gd1}\right)^2}{gm_1^2}\left(\Gamma + \frac{1}{R_f} + \frac{1}{R_2}\right)$$
$$+ \frac{4kTR_1^2\left[1 + \omega^2\left(C_{in} + C_{gs1} + C_{sb2}\right)^2\right]}{\left(1 + gm_1 R_1\right)^2 R_{s1}^2}\left(\Gamma + \frac{1}{R_1}\right) \quad (16)$$

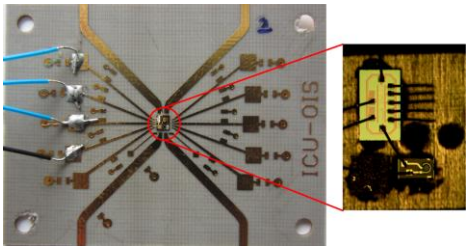
where k is Boltzmann's constant; T is the absolute temperature; Γ is the channel-noise factor of MOSFET; C_{in} is the input parasitic capacitance which includes the photodiode capacitance; bond-pad parasitic capacitance, and electrostatic discharge capacitances ($C_{in}=C_{pd}+C_{ESD}+C_{pad}$). From eq. (16), it can be observed that low frequency noise is dominated by resistor thermal noises and high frequency dominant noise occur due to input parasitic capacitances. The dominant high-frequency noise is divided by $(1+gm_1R_1)$ gain of the local feedback, and hence, the size of local feedback has been increased to reduce total equivalent noises. To reduce the noise generated from the inter-stage, the size of M_3 should be decreased to reduce its gate-source parasitic capacitance, C_{gs3}, and gm_4 should be increased as large as possible. However, increasing the size of M_4 leads to reduction in bandwidth and increase in the input-referred noise. Therefore, the inter-stage transistor sizes have to be chosen carefully not to increase the noise of TIORx or reduce bandwidth performance. Thus, the transconductance of M_3 and M_4 have been chosen to be $gm_3=2gm_4$.

III. Experimental Results

The TIORx was fabricated in a 0.13µm CMOS technology and has a core size (without pads) of 0.051mm^2. The TIORx chip is mounted on wire-bounded chip-on-board (COB) for frequency response, eye-diagram and integrated output noise measurements. The die photograph of TIORx chip and TIORx module are shown in Fig. 2. The TIORx module consists of the TIORx chip, photodiode, and 100 nF single layer capacitor connected between the PD and voltage supply to reduce the ripples from supply voltage. The transimpedance gain (Z_T) of the TIORx chip was calculated from the measured S-parameter data using an Agilent 8703B lightwave component analyzer. Frequency response is measured from 1 GHz to 16 GHz. With a 1KΩ shunt passive feedback, R_F, and 240 fF photodiode capacitance, C_{pd}, a measured transimpedance gain of 60 dBΩ with a 3-dB bandwidth of 6.9 GHz are achieved for the TIORx chip, as shown in Fig. 3.



(a)



(b)

Fig. 2. (a) Die photograph of the fabricated TIORx chip, and (b) TIORx module.

Fig. 4 shows the integrated output noise measured from the output of the TIORx chip with no input connected. The standard deviation of 0.52 mV and by subtracting the oscilloscope noise of 0.1 mV, the corrected integrated noise is 0.42 mV. Fig. 5 shows the graph of input-referred noise based on equation in [8] and measurement data. The estimated noise at 3-dB is 21pA/√Hz.

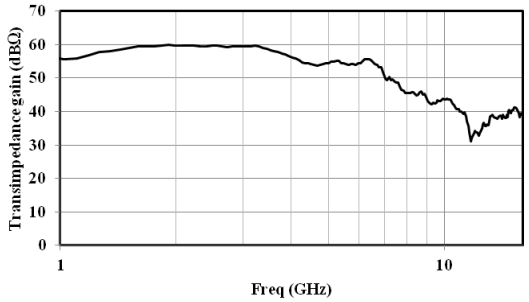


Fig. 3. The measured Z_T (dBΩ) of the proposed TIORx chip.

A 3-dB bandwidth of 6.9 GHz was obtained for the TIORx module as shown in Fig. 6. The eye-diagram have been measured using Anritsu MP1736 pulse-pattern generator with 2^{31}-1 pseudorandom binary sequence input signal, and Agilent 8610A oscilloscope. The eye-diagram is shown in Fig. 7 for the TIORx module. For an applied input light power of -9dBm, an output of 60 mV was achieved for the TIORx module.

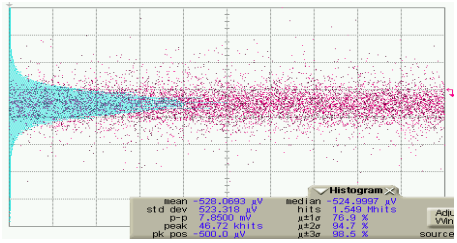


Fig. 4. Measured integrated noise output of the proposed TIORx.

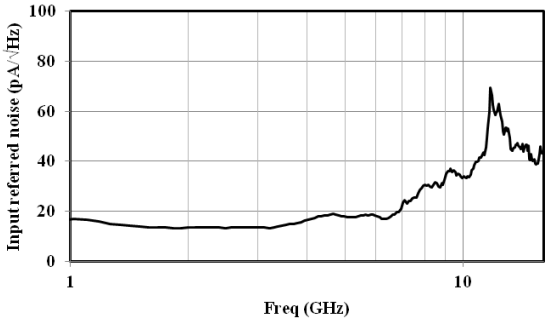


Fig. 5. The measured input-referred noise of the proposed TIORx chip.

The root-mean-sqaure (rms) jitter of 11.57 ps and rise/fall time of 86.7/85.6 ps was obtained (10% to 90%) at 10 Gbps for the TIORx module. The bit-error-rate (BER) performance of the TIORx module has been measured as shown in Fig. 8. A BER performance of less than 10^{-12} was recorded for TIORx module at 10 Gbps with -9 dBm input power. The power dissipation of the TIORx is 16.9 mW at 1.3 V of voltage supply.

TABLE I
COMPARISON OF THE PROPOSED TIA PERFORMANCE WITH OTHER WORKS

Ref.	This work	[8]	[9]	[14]	[15]	[16]
CMOS technology (µm)	0.13	0.13	0.13	0.18	0.18	0.18
Supply	1.3V	1.2V	2V	1.8	1.8	1.8
Power dissipation (mW)	16.9	4.1	98	91.8	70.2	13.97
Inductor	-	Yes	-	Yes	Yes	-
Chip core size(mm^2)	0.051	0.071	0.061	0.8	0.14	-
Data rate (Gbps)	10	10	10	10	10	10
3-dB BW (GHz)	6.9	7.5	6	7.2	7.2	8.5
Gain (dBΩ)	60	50	62	75	61	51.7
Sensitivity 10^{-12} (dBm)	-9	10 µA	22.4 µA	-18.9	10 µA	-
GBP/P_{DC}	408	578	78	441	114	234

Gain bandwidth product per DC power (GHz Ω/P_{DC}), bandwidth (BW).

Table I shows a comparison of the TIORx performance with other works. In our proposed TIORx, a relatively high gain-bandwidth product per DC power (GBP/P_{DC}) of 408 GHzΩ/mW is achieved, compared to the state of the art TIAs. In our proposed design, the addition of inter-stage and post amplifying transconductance stages increases the total noise at higher frequencies. Increase in input-referred noise leads to degradation of input sensitivity. Therefore, the sensitivity in our design is slightly smaller comparing to other artworks. Our proposed TIORx module combines the gain and bandwidth improvement advantages of a conventional Rx module, while acting as a front-end amplifier for low signals from the PD. From Table I, it can also be seen that the size of the proposed TIORx is smaller than the other TIA chips. Thus, the proposed TIORx module consumes low power and occupies a small area when compared to other high data-rate Rx modules of similar CMOS technology.

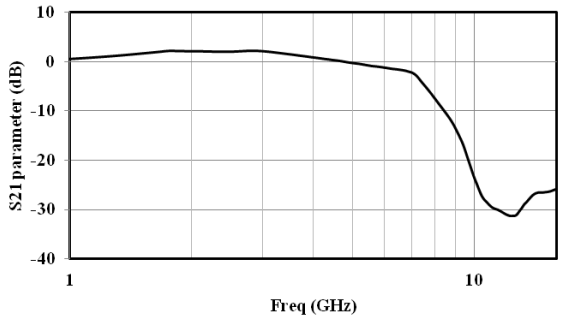


Fig. 6. Normalized S21 measurement result of the proposed TIORx module.

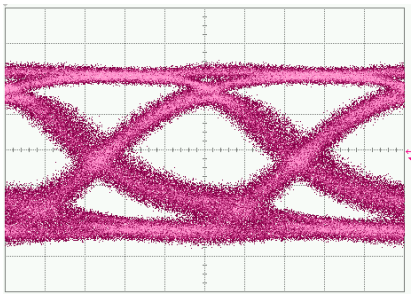


Fig. 7. Eye diagram of the TIORx module with input power of -9 dBm (62µA with 0.5 A/W responsivity of PD), 50 ps/div, 13 mV/div.

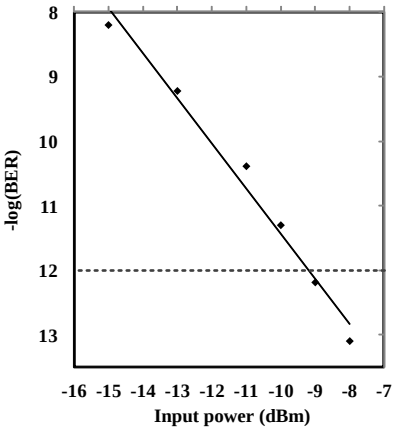


Fig. 8. BER of the proposed TIORx module at 10 Gbps.

IV. CONCLUSION

A TIORx module has been designed and fabricated in 0.13µm CMOS technology for optical interconnect applications which operates up to 10 Gbps data rate. The measured 3-dB bandwidth of the TIORx is 6.9 GHz with a transimpedance gain of 60 dBΩ. While utilizing two intersecting active feedback systems with RGC input stage, the TIORx chip core occupies a small chip area of 0.051mm^2 and has a power consumption of 16.9 mW at 1.3 V. Our proposed TIA-Rx can be applied as a front-end optical Rx to convert the input photo current to output voltage signal, high enough to feed to the next stages such as the De-serializer, PLL, and/or CDR circuits, and it is applicable for chip-to-chip optical interconnects. A GBP/P_{DC} of 408 GHzΩ/mW was achieved. The TIORx module is applicable for chip-to-chip optical interconnects.

REFERENCES

[1] T. Takemoto, F. Yuki, H. Yamashita, Y. Lee, T. Saito, "A compact 4 25-Gb/s 3.0 mW/Gb/s CMOS-based optical receiver for board-to-board interconnects", *J. of Lightwave Technology*, vol. 28, no. 23, Dec. 2010, pp. 3343-3350.

[2] J. Sangirov, I. A. Ukaegbu, T.-W. Lee, M. H. Cho, and H.-H. Park, "Signal Synchronization Using a Flicker Reduction and Denoising Algorithm for Video-Signal Optical Interconnect", *ETRI Journal*, 2012, vol. 34, no. 1, pp. 122-125.

[3] F. Liu, L. Wan, J. Zhou, B. Li, T. Du, W. Gao, F. Wan, "Design optimization and performance verification of multi-channel high-speed optical transceiver package" , *13th Electronics Packaging Technology Conference*, 2011, pp. 153-157.

[4] Zh. Li, et al, "Packaging and assembly of 12-channel parallel optical transceiver module," *Electronic Packaging Technology & High Density Packaging*, pp. 28-30, 2009.

[5] J. Sangirov, A. I. Ukaegbu, T.-W. Lee, M. H. Cho, and H.-H. Park, "Low-power and high-speed SerDes with new dynamic latch and flip-flop for optical interconnect in 180 nm CMOS technology", *Proc.* of *SPIE*, 2011, vol. 7944, pp. 79440V-1-8.

[6] J. Sangirov, I. A. Ukaegbu, Nga T.H. Nguyen, T.-W. Lee, M.-H. Cho, and H.-H. Park, "Comparative Analysis of Single-ended and Differential Receiver Modules in 0.13 µm CMOS technology", IEEE Proc. of ICACT 2013, pp. 217-221.

[7] J. Sangirov, I. A. Ukaegbu, T.-W. Lee, M. H. Cho, and H.-H. Park, "10 Gbps Transimpedance Amplifier-Receiver for Optical Interconnects", *Journal of the Optical Society of Korea*, vol. 17, no. 1, February 2013, pp. 44-49.

[8] T. H. Ngo, T. W. Lee, and H. H. Park, "4.1 mW 50 dBΩ 10 Gbps transimpedance amplifier for optical receivers in 0.13 µm CMOS", *Microwave and Optical Technology Letters*, vol. 53, no. 2, pp. 448-451, Feb. 2011.

[9] O. Momeni, H. Hashemi, and E. Afshari, "A 10-Gb/s inductorless transimpedance amplifier", *IEEE Trans. on Circuits and Systems*, vol. 57, no. 12, pp. 926-930, Dec. 2010.

[10] J. S. Youn, H. S. Kang, M. J. Lee, K. Y. Park, and W. Y. Choi, "10 Gb/s 850-nm CMOS OECC receiver with a silicon avalanche photodetector ", *IEEE Journ. of Quantum Electronics*, vol. 48, no. 2, pp. 229-236, Feb. 2012.

[11] D. Lee, J. W. Han, G. H. Han, and S. M. Park, "An 8.5 Gb/s fully integrated CMOS optoelectronic receiver using slope-detection adaptive equalizer" , vol. 45, no. 12, pp. 2861-2873. Dec. 2010.

[12] S. M. Park and H. J. Yoo, "1.25-Gb/s regulated cascode CMOS transimpedance amplifier for gigabit Ethernet applications", *IEEE Journal of Solid-State Circuits*, vol. 39, no. 1, pp. 112-121, Jan. 2004.

[13] Ch. T. Chan and O. T. Chen, "Inductor-less 10Gb/s CMOS transimpedance amplifier using source-follower regulated cascode and double three-order active feedback", in *Proc. of IEEE Int. Symp. on Circuits and Systems*, 2006, pp. 5487-5490.

[14] J. D. Jin and Sh. S. H. Hsu, "A 75-dBΩ 10-Gb/s transimpedance amplifier in 0.18-lm CMOS technology", *IEEE Photonics Technol Lett*., vol. 20, no. 24, pp. 2177-2179, Dec. 2008.

[15] Ch.-H. Wu, Ch.-H. Lee, W.-Sh. Chen, Sh-I. Liu, "CMOS wideband amplifiers using multiple inductive-series peaking technique" *IEEE Journal of Solid-State Circuits*, vol. 40, no. 2, pp. 548- 552, 2005.

[16] L. Zhenghao, Ch. Dandan, and Y. K. Seng, "An inductor-less broadband design technique for transimpedance amplifiers", in *Proc. of ISIC*, 2009, pp. 232-235.

Jamshid Sangirov received the M.S. degrees in Information and Communication Engineering from Yeungnam University in 2006. He is a Ph.D. student in Electrical Engineering at Korean Advanced Institute of Science and Technology (KAIST) since 2008, he is staying with the Photonic Computer Systems Laboratory, KAIST. He worked with the RFIC Design Team, at Teltron Inc., Korea, from 2010 to 2011. His research interests are analog/RF/VLSI and high-speed electronics, optical interconnections, optical transceiver modules.

Ikechi Augustine Ukaegbu received the B.S. degree in Electrical Engineering, Electro-mechanics, and Electro-technology, the M.S. degree in electronics and microelectronics from Moscow Power Engineering Institute, Technical University, Moscow, Russia, in 2004 and 2006, respectively, and the Ph.D. degree in information and communications engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2012. He was with the System Convergence Technology Team, Convergence and Components Materials Research Lab, Electronics and Telecommunications Research Institute, Korea, from 2008 to 2009. He is currently with the Electrical Engineering Department, Photonics Computer

System Laboratory, KAIST, as a Post-Doctoral Scholar. His current research interests include optical interconnection for high speed and low loss multichannel data links in computer and communication systems, optical interconnections for chip-to-chip and board-to-board data links, chip design for optical transceiver modules, integration, and packaging of optoelectronic components.

Nga T. H. Nguyen received the BS degree in Post Telecommunication Institute of Technology, Hanoi, Vietnam in 2003. She received the M.S degree in August, 2005 and now is Ph.D candidate in Korea Advanced Institute of Science and Technology. Her research interests are low-power and low-cost optical circuit design, and building system for optical interconnection applications.

Tae-Woo Lee received the Ph.D. degree in Electronic and Electrical Engineering from the University of Sheffield, Sheffield, U.K., in 1992. He was an Associate Researcher in the area of compound semiconductor devices with the University of Sheffield, from 1992 to 1994. From 1994 to 2000, he was with the Electronics and Telecommunications Research Institute, Daejeon, Korea, working in the area of high-speed electronic devices and circuits and MMIC. From 2000 to 2003, he was a CEO/CTO with the Venture Company, producing RF PAM and MMIC modules. Since 2003, he has been a Research Professor with the Photonic Systems and Engineering Group, Korea Advanced Institute of Science and Technology, Daejeon. His current research interests include optical interconnections for chip-to-chip and board-to-board data links, optical transceiver modules, and optical switching components for high-speed optical interconnection.

Mu Hee Cho received the Ph.D. received the Ph.D. degree in Physics from Dankook University, Cheonan, Korea, in 1999. He was with the Zenphotonics Company, Daejeon, Korea, from 2000 to 2003, where he was engaged in the research and development of polymer-based optical devices. Since 2003, he has been researching on the chip-to-chip optical interconnections using optical printed circuit boards (PCBs) as a Research Professor with the Photonic Systems and Engineering Group, Korea Advanced Institute of Science and Technology, Daejeon. His current research interests include high-speed optical modules, optical interconnection, and optical PCBs.

Hyo-Hoon Park received the Ph.D. degree in Material Science and Engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 1985. He was a Post-Doctoral Scholar in the area of compound semiconductor devices with Stanford University, Stanford, CA, USA, from 1985 to 1986. From 1986 to 1997, he was with the Electronics and Telecommunications Research Institute, Daejeon, working in the area of high-speed electronic devices and vertical-cavity surface-emitting lasers. Since 1998, he has been a Professor with the Photonic Systems and Engineering Group, KAIST. His current research interests include optical interconnections for chip-to-chip and board-to-board data links, optical transceiver modules, and switching components for optical-link computer systems.

Volume 2 Issue 4, Jul 2013, ISSN: 2288-0003

ICACT-TACT JOURNAL

Global IT Research Institute

1713 Obelisk, 216 Seohyunno, Bundang-gu, Sungnam Kyunggi-do, Republic of Korea 463-824

Business Licence Number : 220-82-07506, Contact: secretariat@icact.org Tel: +82-70-4146-4991