Volunteer Computing: Application for African Scientist

Amoako Gordon Nana Kwesi and Peter Amoako-Yirenkyi

(Invited Paper)

***Abstract*—This paper primarily talks about the art and use of distributed computing specifically in the aspect of volunteer computing and how it can help the African Scientist in the fight against diseases and disasters where high computational simulations are required . A typical case of this is the Africa@Home Project where high demanding computational complexity based research has been implemented using this technology.**

***Index Terms*—Volunteer Computing,BOINC, Africa**

I. INTRODUCTION

VOLUNTEER computing is a technology that takes advantage of a PC's idle time to do meaningful simulations. It takes advantage of your otherwise unused CPU cycles in an effort to do things like predict global climate change, calibrate particle accelerators, or develop drugs to combat cancer and AIDS or malaria. Sometimes it is termed public resource computing.

In volunteer computing, heavy public participation is always available for simulations to be completed..

II. HISTORY OF VOLUNTEER COMPUTING

The term volunteer computing was coined by Luis F.G. Sarmenta, the developer of Bayanihan. The first volunteer computing project on the internet was GIMPS [5],the Great Internet Mersenne Prime Search), which started in January 1996. GIMPS sought to discover new large prime numbers and successfully found 10 to date, including the largest on record, 230,402,457-1. Large prime numbers are very useful tools for data encryption. This was followed by distributed.net [APP3] in 1997 which also solved cryptographic challenges sponsored by RSA Labs and CS Communication & Systems.

Volunteer computing was revolutionalized by the SETI@home[APP4] and Folding@home[APP5] projects which were launched 1999. These projects received considerable media coverage, and each one attracted several hundred of thousands of volunteers.

III. HOW IT WORKS

In volunteer computing, volunteers who are normally general public internet users, willfully provide their computer resources for projects or research. We must reiterate that it relies heavily on public participation. The resources donated by these volunteers are used to do some storage and/or distributed computing geared towards solving a problem. Volunteer computing projects in most cases are for humanitarian logistics research and because of that millions of people worldwide continue to donate their computer resources to help solve problems plaguing the human race.

Volunteer computing projects reside on platforms developed by scientist worldwide. These platforms are lightweight sever/client applications which does the scheduling, distribution of work, security and many other features that handles important routines for any scientist application that demands distributed computing. A typical example of this platform is the open source version called BOINC[APP6].

Volunteers download a client applications BOINC and install. One then attaches himself/herself to an existing project. The volunteer can control the amount of hard disk space or memory to use. When the simulation is done, the result is sent to the central server to be checked with other volunteer results for the same task. When the system verifies the authenticity of the answer, points are awarded to volunteers. The cycle is repeated anytime a volunteer does some work for a particular project. A volunteer can detach himself from a project anytime he/she likes.

The problems solved by volunteer computing projects are ones that could have taken ordinary desktops eternity to solve.

Although Volunteer computing might have security issues because one cannot always know who the client is, it has enormous advantages including but not limited to the fact that it is the cheapest as compared to any form of distributed computing.

IV. ENVIRONMENT FOR AFRICAN SCIENTIST

There are many Projects or organizations who would be more than willing to help African Scientists make use of this facility. One such project is the Africa@Home project.

AFRICA@home is currently a partnership comprising African Scientists, the Swiss Tropical Institute [APP9], the University of Geneva [APP11], the European Organization for Nuclear Research (CERN)[APP7], ICVolunteers[APP2](International Conference Volunteers) and sponsored by Geneva International Academic Network [APP10] . AFRICA@home provides a common framework that involves African students and African universities in the development and running of these volunteer comput-

G.N.K.Amoako is with the Department of Mathematics, Kwame Nkrumah University of Science and Technology,Kumasi,Ghana. Phone mbox+233 20-8127706, e-mail: mboxgordon@math-knust.edu.gh

Peter Amoako-Yirenkyi is also with the Department of Mathematics, Kwame Nkrumah, University of Science and Technology,Kumasi,Ghana. Phone: mbox+233 24-4603503, e-mail: mboxpette@math-knust.edu.gh

ing projects with emphasis on African needs. One success story of Africa@home is the malaria Control[8] project using volunteer computing.

Malaria is responsible for about a million deaths every year in sub-Saharan Africa, and it is the single most biggest killer in children under five. The Malaria Control Volunteer Computing Project[APP8] simulates how malaria spreads through Africa. Running the simulations on thousands of volunteer computers enable researchers to better understand and improve the impact of introducing new treatments.

In a first test phase over few months with 500 volunteers, Africa@Home was able to run simulations equivalent to 150 years of computing time using volunteer computing. The project primarily tries to evaluate the impact of existing and new control interventions, which is different from drug or vaccine development. The main objective initially is to study existing or potential treatments, applied singly or in combination for a range of different settings, primarily for different transmission intensities. The Project actually involves a simulation study that investigates the malaria transmission and clinical epidemiology where the cost effectiveness of simulated interventions are predicted when they are introduced in different combinations. The results from this study is to help planners make decisions on which intervention to choose under which conditions. The more the simulations the better the results and hence the reason for the use of volunteer computing. Results from the first MalariaControl.Net simulation in terms of volunteers computed with an equivalent of 3.0 Teraflops of CPU power are shown in Table 1 below:

	Volunteers	Host PC
Active	4,500	15,000
Total Registered	7,000	20,000

Table 1:*First MalariaControl.Net simulation*

Internet Usage and Population Statistics for Africa for June 30, 2007 [Table 2] showed that, internet usage in Africa constituted only 3% of the world internet usage. Although this is relatively small, African scientists through volunteer computing can access a high percentage of computing resources from the rest of the world and otherwise difficult computing intensive research could easily be solved.

Table 2 shows the distribution of internet usage on the African continent and around the world.

REGION	TFA	RW	Total
POP	933,448,292	5,641,218,125	6,574,666,417
PIW	14.2%	85.5%	100%
IULD	43,995,700	120,453,901	1,244,449,601
P%P	4.7%	21.3%	18.9%
UIW	3.5%	96.5%	100%
UG	874.6%	236.8%	244.7%

Table 2:*Internet users and population statistics for Africa.*

POP: Percentage of estimated Population(2007 Est.), PIW: Percentage of Population in World, IULD Internet Users Latest Data, P%P: Penetration of Population), UIW: Percentage of Users in World UG: User Growth (2000-2007). TFA: Total for Africa, RW: Rest of the world and WT: World Total.

BOINC can automatically match work to be processed with hosts suitable to execute it, taking into account estimated memory and disk requirements as well as architecture and operating system constraints.

V. BUILDING A BOINC PROJECT

In this paper we assume the BOINC server is setup and fully running. For more information about installing a BOINC server, please visit the website at http://boinc.berkeley.edu/.

After installation, for the scientist/researcher to run his 'huge-draining' application on the BOINC platform, he first has to make a project. The make_project script creates the server components of a BOINC project. The link to making a BOINC project is http://boinc.berkeley.edu/trac/wiki/MakeProject.

There are quiet a number of ways a scientist or a researcher can make use of the platform after the project has been made. The application can be built to run on a lot of architectures or platforms, but the preference is left to the scientist.

David Anderson, the main brain behind the BOINC application in his paper[2] described a boinc project as a project that corresponds to an organization or a research group that does public-resource computing. A project is identified by a single master URL, which is the home page of its web site and also serves as a directory of scheduling servers. volunteers register with projects. A project can involve one or more applications, and the set of applications can change over time.

Every BOINC Project has a relational database that stores descriptions of applications, platforms, versions, workunits, results, accounts, teams, and so on. Every BOINC Server that manages a Project comprises of several functions executed by a set of web services and daemon processes. Among these are Scheduling and Data or File transfer servers. The Scheduling servers handles RPCs from clients by issuing work and handling reports of completed results. Data servers handles file uploads using certificate-based mechanism to ensure that only legitimate files, with prescribed size limits, can be uploaded. File downloads are handled by plain HTTP. BOINC also makes it possible to use Python scripts and C++ interfaces for creating, starting, stopping and querying projections, creating workunits, and monitoring server performance.

Volunteer Computing provides tools that let volunteers remotely install the client software on large numbers of machines, and attach the client to accounts on multiple projects. Volunteers or participants can join a BOINC-based project by visiting the projects web site, filling out a registration form, and downloading the BOINC client. The BOINC client could be run as a windows service which logs errors to a database even when no users are logged

in. It has a tabular view that displays projects, work, file transfers, and disk usage. The client may decide to just use a screensaver option showing the running applications. In UNIX it can be run as a command-line program that communicates through stdin, stdout and stderr from and can be run from a cron job or startup file.

VI. CONCLUSION

There is an increasingly growth factor in internet usage worldwide and that is in favour of volunteer computing. The power of this technology could therefore be harnessed to solve scientific problems on the African Continent. The onus lies on the African scientist to make good use of a simple but yet powerful technology in solving useful problems and Africa would see an overwhelming improvement in the control of malaria and other deadly diseases. And most importantly Africa can rely on the world's internet community to solve problems once they are able to think through scientifically.

VII. ACKNOWLEDGMENT

We would like to thank Viola Krebs and the ICVolunteers team for organizing a great workshop which was an eye-opener, to all the tutors who taught us the concept of volunteer computing, Nicolas Maire(STI), Joris Borgdoff(Erasmus MC, Dept. of Public Health), Chris Sutton, Daniel Lombraña González(University of Extremadura, Spain) and Christian Ulrik Sttrup(Niels Bohr Institute).

The authors are also grateful to Ben Segal and Francois Grey (CERN) for inspiring us. To David Anderson and his team, we say congratulations.

APPENDIX

[APP1] Bayanihan is an old tradition in the Philippines where neighbors of a relocating family would help the family move by gathering under their house and carrying it to its new location. Although Bayanihan practiced in this form has become rare in today's modern times, the word bayanihan itself has come to mean any manifestation of the powerful spirit of communal unity that can make seemingly impossible feats possible through the cooperation of many people working towards a common goal. The bayanihan spirit employed in the realm of computing makes it possible and easy for people to pool together their computing resources into a high-performance computing resource capable of solving computational problems much faster than before.

[APP2] ICVolunteers is an international non-profit organization specialized in the field of communications, in particular languages, conference support and cybervolunteerism. They work with volunteers to implement social and educational programs in order to help populations and local communities to develop. Through volunteer effort, they cooperate with organizations in the humanitarian, social, environmental and medical fields to implement projects and conferences at local, national and international levels. In addition, ICVolunteers promotes volunteerism and its recognition, by enhancing civic commitment and involvement, and by providing leadership and links between organizations, individuals and communities. With headquarters in Geneva (Switzerland), ICVolunteers has offices and permanent representation in a number of other countries, including France, South Africa, Mali and Spain. [APP3] distributed.net (or Distributed Computing Technologies, Inc. or DCTI) is a world-wide distributed computing effort that is attempting to solve large scale problems using otherwise idle CPU time. It is officially recognized as a non-profit organization under U.S. tax code 501(c)(3)

[APP4] SETI@home ("SETI at home") is a distributed computing project using Internet-connected computers, hosted by the Space Sciences Laboratory, at the University of California, Berkeley, in the United States. SETI is an acronym for the Search for Extra-Terrestrial Intelligence. SETI@home was released to the public on May 17, 1999.[4][5][6]

[APP5] Folding@Home - assesses protein structures to find cures to serious diseases.

[APP6] The Berkeley Open Infrastructure for Network Computing (BOINC) is a non-commercial middleware system for volunteer computing. It is useful for legacy applications in areas as diverse as mathematics, medicine, molecular biology, climatology, and astrophysics. Visit http://boinc.berkeley.edu/

[APP7] CERN, the European Organization for Nuclear Research, is the world's leading laboratory for particle physics. It has its headquarters in Geneva. At present, its Member States are Austria, Belgium, Bulgaria, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Italy, Netherlands, Norway, Poland, Portugal, Slovakia, Spain, Sweden, Switzerland and the United Kingdom. India, Israel, Japan, the Russian Federation, the United States of America, Turkey, the European Commission and UNESCO have Observer status.

[APP8] MalariaControl.net uses the same BOINC software platform that allows hundreds of thousands of people worldwide to participate in projects such as SETI@home and Climateprediction.net, searching for signs of extraterrestrial intelligence or forecasting the climate in the 21st century.

[APP9] The Swiss Tropical Institute (STI)is based in Basel but has activities worldwide in support of its mandate to contribute to the improvement of the health of populations internationally and nationally through excellence in research, services, and teaching and training. It is a statutory organization with core support from the Swiss Federal Government and the Canton of Basel-Stadt. The malaria modeling activities are supported by the Bill & Melinda Gates Foundation.

[APP10] The Geneva International Academic Network (GIAN) is an international research network whose primary objective is to reinforce cooperation among in-

ternational organizations and academic institutions. The GIAN funds research activities that involve a partnership between the academic world and international organizations and that concern at least one of five thematic areas: globalisation, sustainable development, social equity, intercultural dialogue or human rights. The GIAN benefits from the collaborative and financial support of the Swiss Confederation and the Republic and Canton of Geneva.

[APP11] University of Geneva is the second largest university in Switzerland and is a public institution of the Republic and Canton of Geneva. It pursues three missions: teaching, research, and service to the wider community. From the time of its creation in 1559 by Jean Calvin, right up to the recent discovery by University astrophysicists of extrasolar planets, the University of Geneva has continued to grow and develop while maintaining its longstanding tradition of excellence with an international angle.

References

[1] D. Kondo, D.P. Anderson, and J. McLeod VII. "Performance Evaluation of Scheduling Policies for Volunteer Computing", 2007
[2] David P. Anderson, *BOINC:A system for public resource Computing, 5th IEEE/ACM International Workshop on Grid Computing, November 8, 2004, Pittsburgh, USA, pp 1-7.*
[3] Carl Christensen, Tolu Aina and David Stainforth *The Challenge of Volunteer Computing with Lengthy Climate Model Simulations*
[4] http://boinc.berkeley.edu/, 2007
[5] http://en.wikipedia.org/wiki/Volunteer_computing, 2007
[6] http://press.web.cern.ch/press/PressReleases/Releases2006/PR07.06E.html, 2007
[7] http://www.internetworldstats.com/stats1.htm, 2007
[8] http://en.wikipedia.org/wiki/Distributed.net, 2007