

EXPLORING GRID COMPUTING & VOLUNTEER COMPUTING: ANALYZING DAILY COMPUTING RUNTIMES ON THE WORLD COMMUNITY GRID

*Kieran Raabe, Bloomsburg University of Pennsylvania, kjr10324@huskies.bloomu.edu
Loreen M. Powell, Bloomsburg University of Pennsylvania, lpowell@bloomu.edu*

ABSTRACT

In the early 1990s, the concept of grid computing was little more than a simple comparison which stated that making computing power as accessible as electricity from a power grid. Today, grid computing is far more complex. Grid computing involves utilizing unused computing power from devices across the globe to run calculations and simulations and submit the results to move a project closer to its goal. Some projects, such as World Community Grid, work to achieve numerous goals, herein referred to as subprojects. The goal of this research is to explore grid computing and volunteer computing. Specifically, this research involved a daily collection of statistics on daily results returned to World Community Grid and daily computing runtimes for three subprojects named Africa Rainfall Project, Microbiome Immunity Project, and Help Stop TB. The data collection lasted four weeks and resulted in five graphs for each subproject being created with the collected data. Results revealed a correlation between daily results returned and daily runtimes, with each data point for results returned being slightly lower than the runtime data point from the same day.

Keywords: Grid Computing, Volunteer Computing, BOINC, World Community Grid

INTRODUCTION

The concept of grid computing has existed since the early 1990s. Originally “a metaphor for making computer power as easy to access by the user as electricity from a power grid” (Dutton & Jeffreys, 2010). Today, grid computing is the process of setting up a network of computers and, in some cases, devices as small as smartphones and tapping into the currently unused resources these computers can provide. Hence, grid computing architecture, resource availability and scalability are far more complex and geographically dispersed (Wided & Okba, 2019). As result, grid computing has the potential to have a profound impact on the world. In fact, grid computing makes volunteer computing possible. Volunteer Computing harnesses consumer digital devices to provide computing infrastructure for scientific experiments, big data collection and analysis (Anderson, 2019; Mengistu & Che, 2019). However, grid computing and volunteer computing are rarely explained or discussed in introductory or principles of computing courses.

In an effort to provide a better understanding of grid computing and extend the learning experience beyond the walls of the classroom (Beck, 2015; Levin-Goldberg, 2012), this paper explores current grid computing literature. Additionally, this research uses the World Community Grid to analyze daily computing runtimes on the world community grid. Specifically, this research involved a daily collection of statistics on daily results returned to World Community Grid and daily computing runtimes for three subprojects named Africa Rainfall Project, Microbiome Immunity Project, and Help Stop TB. This work has practical implications for IT programs by providing a resource for faculty and students wishing to better understand grid computing. This research could be replicated at other institutions. The remainder of this paper is structured as follows: review of literature, methodology, results, and conclusion.

LITERATURE REVIEW

Grid Computing

There are many large tasks that exist that require an extensive amount of computational resources and time to complete them even with supercomputers. One possible solution is to utilize grid computing. Grid computing allows for a queuing systemic process to allocate resources to processors for a certain interval where they once remained idle (Afanasyev, Bychkov, Zaikin, Manzyuk, Posypkin, & Semenov, 2017). Hence, the unused computer resources, such as processing power, memory, and even the computers' graphics processing unit, can then be utilized to run tests and simulations in a collective effort to complete extensive tasks.

The grid computing concept originated in the early to mid-1990s working only with supercomputers for compute intensive tasks (Janciak, Brezany, Tjoa, 2004; Foster & Kesselman, 1999). Over the years, grid computing has expanded into so much more beyond the use of just supercomputers and compute intensive tasks to also focus on data-intensive applications and advanced scientific research projects.

Volunteer Computing

Volunteer computing is a form of grid computing. Contributors use their devices' unused processing power, memory, or, in some cases, graphics processing unit to conduct tests and simulations and return the results to the project (Anderson, 2019). The idea behind volunteer computing is that participating computers donate computing time and resources during their idle times. For example, many university laboratory computers that are not in use or don't have a large portion of computing resources used can be volunteer computers. Additionally, many companies allocate computers during employees' lunch, breaks, or out of the office times to participate as a volunteer computer (Kijispongse, Piyatumrong & U-ruekolan, 2018). Hence, volunteer computing is a novel way for many organizations and individuals to donate the electronic device(s) to help solve, aid, or enhance scientific projects.

To participate in volunteer computing, one typically registers with a volunteer project. Client software must be downloaded and installed on the participating electronic device. The client software monitors the participating computer for an idle or high available computing resource state to obtain a task from the server. The participating computer will perform the task and return the completed task to the server (Kijispongse et. al., 2018).

It is important to note that volunteer computing systems success depends on the task distribution and scheduling (Mengistu & Che, 2019). Nouman, Durrani and Shamsi (2014) classify task distribution policies into naïve scheduling or knowledge based. In a naïve scheduling policy, tasks are not distributed based upon volunteer history. Whereas knowledge-based policies are distributed based upon volunteer history.

A recent article by Mengistu and Che (2019) provided a comprehensive review of a volunteer computing system's technical requirements of each type and classification. They stated that regardless of the classified task policy, tasks are often divided into subtasks. The subtasks can help volunteer nodes perform efficiently and optimize scheduling. However, they argue that regardless of the subtasks, knowledge-based policies help to ensure that tasks are completed within the time constraint because they can rely on the volunteer nodes past history. Hence, making knowledge-based policies more attractive for large reliable projects like the World Community Grid (Estrada et al. 2008; Mengistu & Che, 2019; Toth 2008; Toth & Finkel 2009).

The World Community Grid (<http://www.worldcommunitygrid.org>) is a project that focuses on environmental and health related research, relies heavily on volunteer computing for its subprojects. Basically, anyone with a smart electronic device can allocate their idle computing power to progress scientific research. A few projects that the World Community Grid offers are the Africa Rainfall Project, the Microbiome Immunity Project, and the Help Stop TB project. It is estimated that "750,000 individuals and 470 organizations from 80 countries have helped World Community Grid on 26 research projects" (Taft, 2016).

Volunteer computing is not only a great way for a user to contribute to the scientific community. This offshoot of grid computing can also pave the way for a data crunching powerhouse due to the sheer number of FLOPS, or floating point operations per second, every device contributes. For instance, on May 26, 2020, Berkeley Open Infrastructure for Network Computing (BOINC) grid's virtual supercomputer clocked an average of 32,743.827 TeraFLOPS. This

massive number of FLOPS results in far superior performance compared to, for example, the Xbox Series X custom processor, which “Microsoft rates at 12 [TeraFLOPS]” (Martindale, 2020) and the Mac Pro that “can reach up to 56 [TeraFLOPS] of power” (Martindale, 2020). With the performance of approximately 2,728 of the Xbox Series X custom processors, researchers dependent on volunteer computing can expect swift contributions towards their main goals.

Berkeley Open Infrastructure for Network Computing (BOINC)

Grid computing needs to occur in an organized environment to be effective (Afanasiev, et. al., 2017). These environments are called grids. Grids, such as Berkeley Open Infrastructure for Network Computing (BOINC, <https://boinc.berkeley.edu/>), are comprised of many less powerful devices which are connected to a virtual supercomputer. BOINC is a commonly used open-source middleware system for grid computing and volunteer computing (Anderson, 2004; Anderson, 2019).

Afanasiev, et al. (2017) lists five basic principles of BOINC. They include an administration center, project resources, BOINC manager at the host, intermediate results are saved at the host, and priority multitasking of the BOINC manager. Ivashko and Nikitina (2013) and Afanasiev et al (2017) believe that BOINC is one of the most popular platforms for volunteer computing projects. Currently, there are thirty different projects that utilize BOINC for scientific research (University of California, 2020). One interesting scientific project utilizing BOINC is World Community Grid.

HYPOTHESIS

The goal of this research is to explore grid and volunteer computing on BOINC by analyzing daily computing runtimes on the world community grid. Our research centered upon the following research question: Is Grid computing at an optimal state?

The research hypothesis to be tested is as follows:

H₁: As daily results returned totals and averages increase, daily runtime totals and averages will also increase, but at a slower pace.

If the hypothesis is proved, then the concept of grid computing will be considered an optimal, state of the art method of completing research and achieving world-changing milestones. If the hypothesis was disproved, then grid computing may require more time to develop and fine-tune its capabilities to better achieve these lofty expectations. Note that, even if the hypothesis is disproved, grid computing remains an efficient method of research contribution and its power should not be taken lightly.

One variable and potential threat to the outcome of this research is use of cellular devices. The BOINC grid has a mobile application that allows users to contribute to projects like World Community Grid. Since cell phones tend to have a lower performance compared to desktop computers or laptops, the use of cell phones may result in increased daily runtimes and, consequently, a decrease in daily results returned. Hence, a sub question to be examined is the following:

SQ₁: Will cell phone data increase daily runtimes and decrease the daily results returned?

RESEARCH METHODOLOGY

Data was collected over a four-week period. Specifically, data was collected between April 1, 2020 and April 28, 2020. Four types of data were collected: average runtime per day, total runtime per day, average daily results returned, and total daily results returned. This data was recorded from three of World Community Grid’s subprojects: the Africa Rainfall Project, the Microbiome Immunity Project, and the Help Stop TB project. These specific subprojects were selected because each subproject was in a different life stage, with the Africa Rainfall Project being new and, on the rise, the Microbiome Immunity Project operating at peak performance, and the Help Stop TB project nearing completion and slowing down as a result.

Additionally, the authors selected to gather data from the Africa Rainfall Project since this was the World Community Grid's latest project. Thus, the researchers believed that the World Community Grid users would begin contributing in masses in order to give this project a good head start.

Microsoft Excel was used to record the data and graph the results. Line and stacked line charts were selected as the best way to present the results. Figure 1 displays the process diagram which describes how the data used in this research was retrieved.

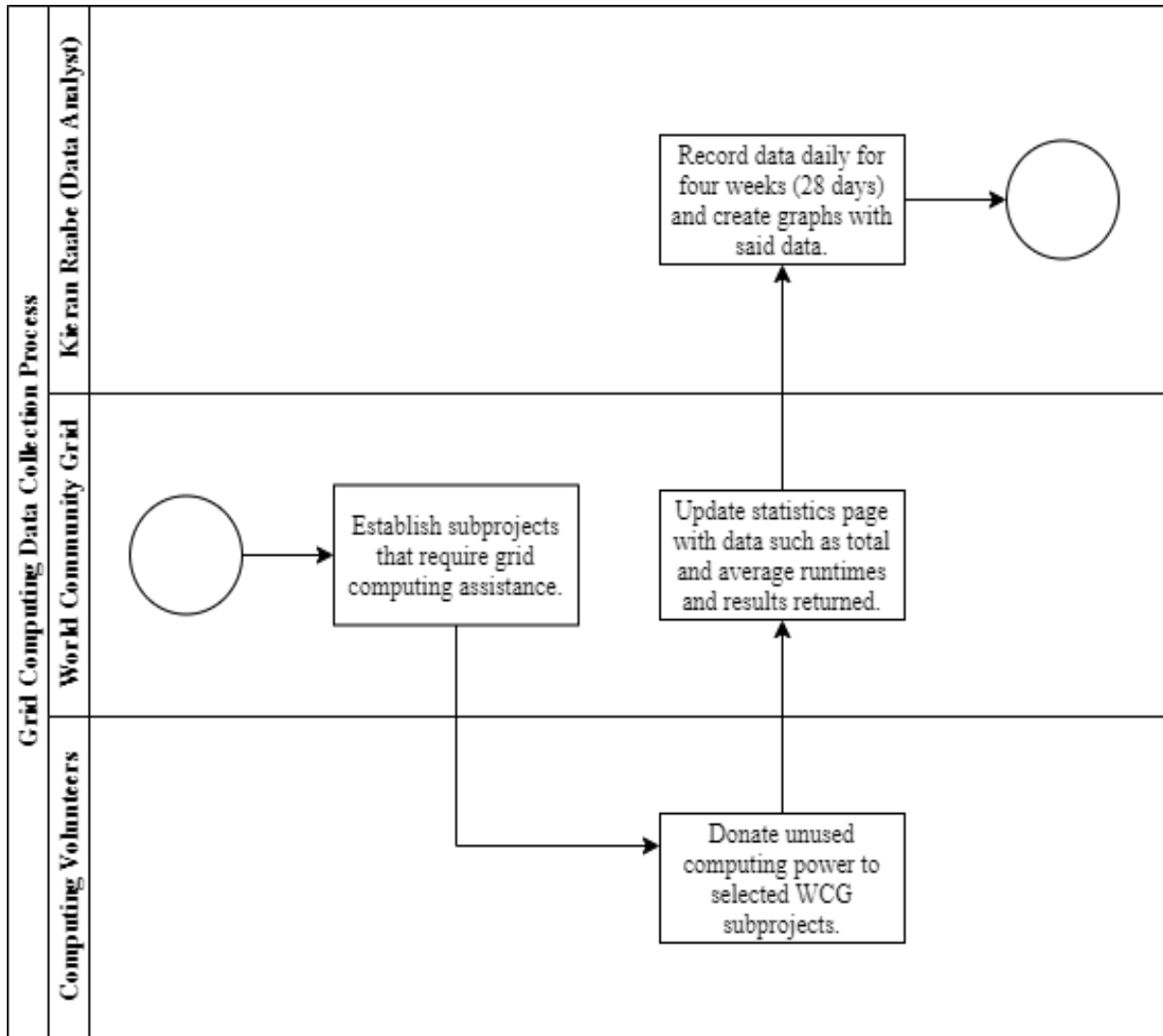


Figure 1: The Research Process Diagram

RESULTS

This section presents results that were obtained by the data collected. Figure 2 displays the BOINC statistics on May 26, 2020, which includes the average FLOPS. Additionally, tables 1, 2, and 3 display the individual types of collected data, along with the raw data, for each of the three subprojects.

First seen on	2004-07-30 12:30:01
Category	Uncategorized
Total credit	38,997,518,310.093 Comprising 38,994,213,681,948 from the daily update + 3,304,628,144 since then
Recent average credit RAC	6,548,765,405
Recent average credit RAC (according to BOINCstats)	6,342,942,071
Average floating point operations per second	32,743,827.0 GigaFLOPS / 32,743.827 TeraFLOPS
Users	4,012,927
Active users	147,726 (3.68%)
Hosts	987,677
Active hosts	794,334 (80.42%)
Teams	110,262
Active teams	14,862 (13.48%)

Figure 2: BOINC's Statistics on May 26, 2020, Including Average FLOPS.

Table 1: The Individual Types of Collected Data, Along with the Raw Data, for the Africa Rainfall Project.

Africa Rainfall Project	Day	Avg. Runtime/Calendar Day	Total Daily Runtime	Avg. Daily Results Returned	Total Daily Results Returned
	4/1/2020	9.073972603	21.6	3,250.34	8,217
	4/2/2020	9.136986301	20.24931507	3,273.25	7,450
	4/3/2020	9.208219178	18.13972603	3,297.94	6,615
	4/4/2020	9.276712329	21.34246575	3,320.09	7,181
	4/5/2020	9.342465753	19.99726027	3,341.28	6,858
	4/6/2020	9.367123288	16.63835616	3,348.54	5,622
	4/7/2020	9.4	13.31232877	3,361.02	4,650
	4/8/2020	9.424657534	13.84931507	3,372.30	5,312
	4/9/2020	9.44109589	13.31780822	3,380.55	5,052
	4/10/2020	9.465753425	12.54794521	3,392.51	4,922
	4/11/2020	9.498630137	13.83561644	3,406.91	5,535
	4/12/2020	9.545205479	15.50684932	3,424.40	5,943
	4/13/2020	9.610958904	18.8739726	3,445.02	6,605
	4/14/2020	9.687671233	22.42465753	3,470.70	7,471
	4/15/2020	9.756164384	21.53972603	3,495.87	7,704
	4/16/2020	9.81369863	20.34794521	3,519.19	7,535
	4/17/2020	9.873972603	19.98356164	3,541.74	7,560
	4/18/2020	9.947945205	21.6109589	3,568.27	7,680
	4/19/2020	10.02739726	22.59726027	3,594.97	8,179
	4/20/2020	10.09863014	23.71506849	3,621.37	8,293
	4/21/2020	10.16712329	21.81643836	3,649.22	8,191
	4/22/2020	10.22739726	21.49863014	3,674.09	8,307
	4/23/2020	10.27945205	17.89041096	3,694.93	6,826
	4/24/2020	10.30684932	19.84109589	3,706.61	7,512
	4/25/2020	10.36986301	15.3260274	3,734.24	6,002
	4/26/2020	10.41917808	18.29589041	3,754.77	7,092
	4/27/2020	10.46849315	19.15068493	3,775.85	7,508
	4/28/2020	10.71320755	20.14716981	3,796.85	7,807

Table 2: The Individual Types of Collected Data, Along with the Raw Data, for the Microbiome Immunity Project.

Microbiome Immunity Project	Day	Avg. Runtime/Calendar Day	Total Daily Runtime	Avg. Daily Results Returned	Total Daily Results Returned
	4/1/2020	118.0383562	137.5945205	505,443.80	674,492
	4/2/2020	118.0630137	138.7452055	505,648.18	680,773
	4/3/2020	118.0849315	145.569863	505,874.66	726,569
	4/4/2020	118.1041096	132.5863014	506,048.29	680,393
	4/5/2020	118.1369863	142.3863014	506,296.89	708,615
	4/6/2020	118.169863	154.2410959	506,556.10	775,690
	4/7/2020	118.1835616	138.9808219	506,734.12	700,054
	4/8/2020	118.1917808	127.1835616	506,885.20	663,944
	4/9/2020	118.2027397	128.1589041	507,034.07	648,624
	4/10/2020	118.2109589	128.9726027	507,148.32	641,871
	4/11/2020	118.2191781	126.2246575	507,256.50	601,730
	4/12/2020	118.2219178	122.1068493	507,376.29	626,324
	4/13/2020	118.230137	123.2136986	507,496.70	617,665
	4/14/2020	118.2356164	123.9452055	507,600.10	617,289
	4/15/2020	118.2410959	123.1753425	507,691.92	598,928
	4/16/2020	118.2465753	122.9534247	507,782.57	597,705
	4/17/2020	118.2520548	126.139726	507,888.51	605,985
	4/18/2020	118.2547945	121.2465753	507,974.28	599,716
	4/19/2020	118.260274	121.0438356	508,062.34	590,862
	4/20/2020	118.2684932	124.490411	508,148.32	594,702
	4/21/2020	118.2876712	133.1123288	508,211.59	578,534
	4/22/2020	118.3068493	136.1205479	508,271.49	567,506
	4/23/2020	118.3260274	115.3506849	508,323.31	469,722
	4/24/2020	118.3561644	167.4164384	508,399.24	671,260
	4/25/2020	118.4	140.4657534	508,520.01	568,078
	4/26/2020	118.4246575	139.4547945	508,591.35	571,759
	4/27/2020	118.4493151	143.2739726	508,662.26	580,627
	4/28/2020	118.6603774	143.7945205	508,755.81	586,886

Table 3: The Individual Types of Collected Data, Along with the Raw Data, for the Help Stop TB project.

Help Stop TB	Day	Avg. Runtime/Calendar Day	Total Daily Runtime	Avg. Daily Results Returned	Total Daily Results Returned
	4/1/2020	4.369863014	1.156164384	3,215.17	610
	4/2/2020	4.367123288	1.367123288	3,213.46	768
	4/3/2020	4.364383562	1.202739726	3,211.78	672
	4/4/2020	4.361643836	1.364383562	3,210.12	743
	4/5/2020	4.361643836	1.734246575	3,208.55	877
	4/6/2020	4.35890411	1.8	3,207.02	973
	4/7/2020	4.356164384	1.553424658	3,205.33	787
	4/8/2020	4.356164384	1.246575342	3,203.51	595
	4/9/2020	4.353424658	1.076712329	3,201.76	543
	4/10/2020	4.350684932	1.139726027	3,199.91	550
	4/11/2020	4.347945205	1.01369863	3,198.12	510
	4/12/2020	4.345205479	1.136986301	3,196.30	580
	4/13/2020	4.342465753	0.81369863	3,194.39	402
	4/14/2020	4.342465753	0.783561644	3,192.55	389
	4/15/2020	4.339726027	1.230136986	3,190.91	648
	4/16/2020	4.336986301	1.282191781	3,189.16	666
	4/17/2020	4.334246575	1.375342466	3,187.60	738
	4/18/2020	4.334246575	1.580821918	3,185.99	849
	4/19/2020	4.331506849	1.268493151	3,184.20	616
	4/20/2020	4.328767123	1.167123288	3,182.48	609
	4/21/2020	4.326027397	1.249315068	3,180.78	615
	4/22/2020	4.326027397	1.394520548	3,179.17	698
	4/23/2020	4.323287671	1.265753425	3,177.51	674
	4/24/2020	4.320547945	1.243835616	3,175.66	630
	4/25/2020	4.317808219	1.109589041	3,172.19	627
	4/26/2020	4.315068493	1.221917808	3,170.39	670
	4/27/2020	4.312328767	0.824657534	3,168.54	434
	4/28/2020	4.309589041	0.769863014	3,166.77	414

Figures 3, 4, and 5 are three comparison graphs which display the total daily runtimes for each project at a higher point than the total daily results returned. This outcome was the exact opposite of the hypothesis, therefore disproving it. Originally, the use of less powerful mobile devices was thought to hold the most blame. However, after taking a closer look, something seemed to even disprove that thought.

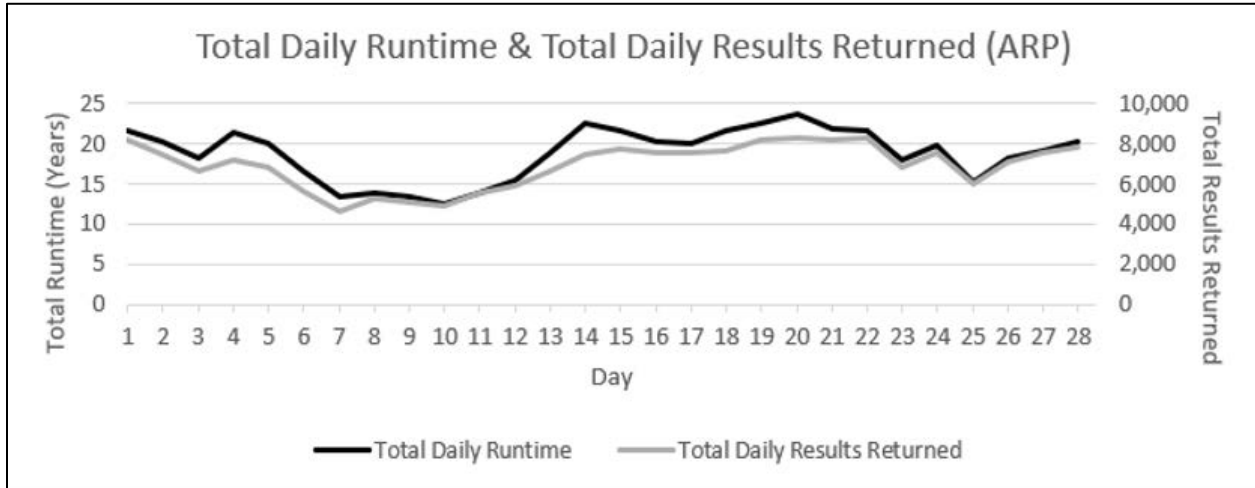


Figure 3: Comparison of Total Daily Runtime and Results Returned Data Points for Africa Rainfall Project

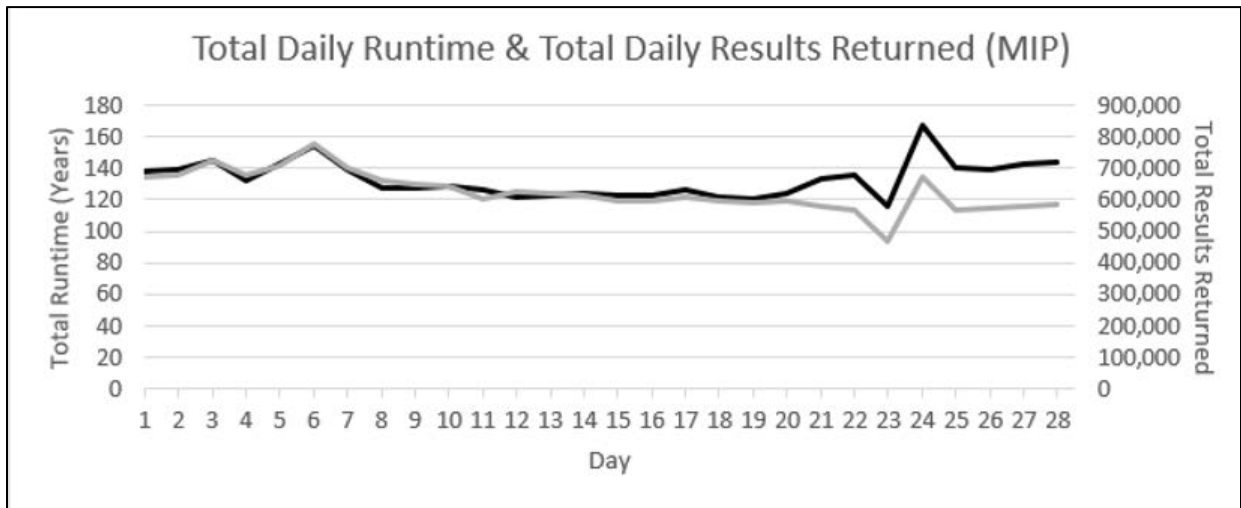


Figure 4. Comparison of Total Daily Runtime and Results Returned Data Points for Microbiome Immunity Project

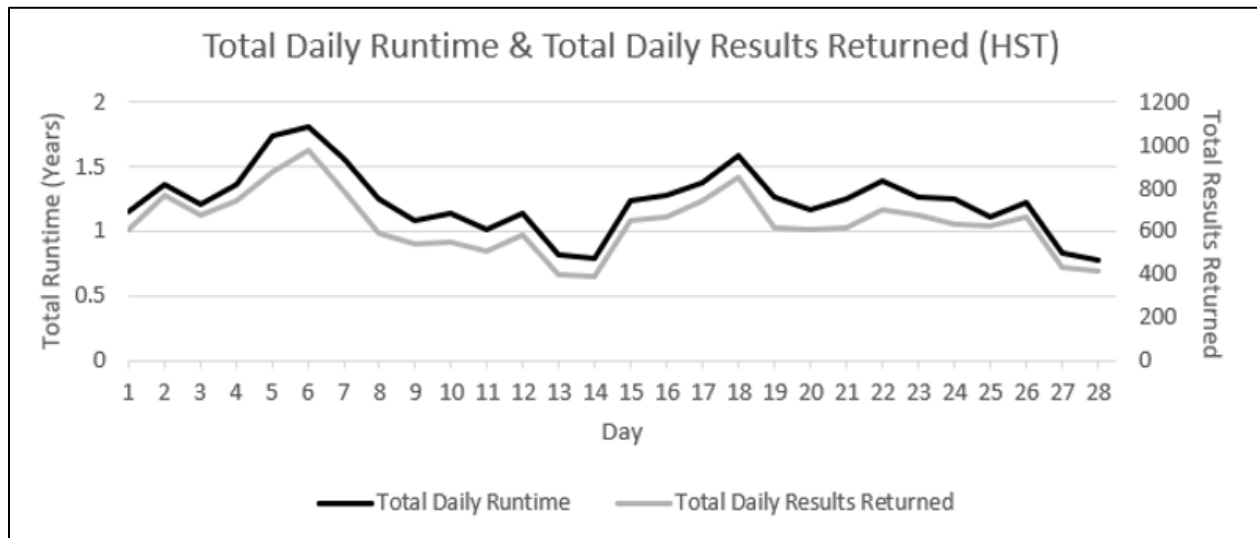


Figure 5. Comparison of Total Daily Runtime and Results Returned Data Points for Help Stop TB Project

As listed above in figures 3, 4, and 5, the lines on the comparison graphs were increasing and decreasing in nearly perfect tandem. The outcome was that as the total daily runtimes increased, the total daily results returned increased by close to the same magnitude. A prime example is shown in the comparison graph for the Microbiome Immunity Project, figure 3, on day 24. Though the total daily results returned was at a lower point than the total daily runtime, both types of data increased at almost exactly the same rate from the previous day, with the total daily results returned increasing by 42.9% and the total daily runtime increasing by 45.1%. In other words, it does not appear to make a large difference whether mobile devices are used to contribute to these subprojects, and it is unclear if mobile devices are commonly used to contribute at all.

CONCLUSION

Although the original hypothesis was disproven, the strikingly similar increasing and decreasing rates of the total daily results returned and the total daily runtimes was surprising. This work provided practical implications for IT programs by providing a resource for faculty and students wishing to better understand grid and volunteer computing.

This research is not without limitations. First, the research is limited to the World Community Grid's subprojects: the Africa Rainfall Project, the Microbiome Immunity Project, and the Help Stop TB project. Second, it is also limited in the time frame of data collection. Future research should address these limitations and continue to expand upon this study.

REFERENCES

- Anderson, D. P. (2019). *BOINC: A Platform for Volunteer Computing*. Accessed <https://arxiv.org/abs/1903.01699>
- Anderson, D.P. (2004) BOINC: a system for public-resource computing and storage. In: *Proceedings of the 5th IEEE/ACM International Workshop on Grid Computing (Grid)*, pp 4–10.
- Afanasiev, P., Bychkov, I.V., Zaikin, O.S., Manzyuk, M.O., Posypkin, M.A., & Semenov, A.A. (2017). Concept of a multitask grid system with a flexible allocation of idle computational resources of supercomputers. *Journal of Computer and Systems Sciences International*, 56(4), 701–707.
- Dutton, W. H., & Jeffrey, P. W. (2010). *Worldwide research: reshaping the sciences and humanities*. Cambridge, MA, MIT Press.
- Foster, I. & Kesselman, C. (1999). *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann, San Francisco, CA.

- IBM: World Community Grid participation. (2017). *Healthcare Leadership Review*, 36(2), 17.
- Ingram, A., Leverette, J., & Nguyen, K. (2019). Alabama power's smart neighborhood: Behind the scenes of the southeast's first community-scale microgrid, which brings value to both the homeowners and the grid. *T&D World*, 1.
- Ivashko, E. E & Nikitina, N. N. (2013). Use of BOINC-grid in computationally intensive scientific studies, Vest. Novosib. Univ., *Service Information Technology*. 11, 53–57.
- Janciak I., Brezany P., Tjoa A.M. (2004). Towards the wisdom grid: goals and architecture. In: Wyrzykowski R., Dongarra J., Paprzycki M., Waśniewski J. (eds) *Parallel Processing and Applied Mathematics. PPAM 2003. Lecture Notes in Computer Science*, vol 3019. Springer, Berlin, Heidelberg, 796-803.
- Jianhua YU, & Yuan LI. (2018). New methods of un-cheatable grid computing. *Computing & Informatics*, 37(6), 1293–1312.
- Kijsipongse, E., Piyatumrong, A. & U-ruekolan, S. (2018). A hybrid GPU cluster and volunteer computing platform for scalable deep learning. *Journal of Supercomputing*, (2018)74, 3236–3263
- Martindale, J. (2020). *What Is a Teraflop?*. Accessed <https://www.digitaltrends.com/computing/what-is-a-teraflop/>
- Meddeber, M., & Yagoubi, B. (2019). Dependent tasks assignment and data consistency management for grid computing. *Multiagent & Grid Systems*, 15(2), 179–196.
- Mengistu, T. M., & Che., D. (2019). Survey and taxonomy of volunteer computing. *ACM Computing Surveys*, 52(3), 1–35.
- Taft, D. K. (2016). IBM enlists world community grid to combat tuberculosis. *EWeek*, 1.
- Wided, A., & Okba, K. (2019). A novel agent based load balancing model for maximizing resource utilization in grid computing. *Informatica*, 43(3), 355–361.