The GO Competition Sandbox

Stephen Elbert, Pacific Northwest National Laboratory

Challenge Background

ARPA-E's Grid Optimization (GO) Competition (https://gocompetition.energy.gov) seeks to improve the optimization algorithms that control the operation of the U.S. Electric Grid to reduce the expense of the U.S. energy sector while improving reliability. The 2013 FERC Technical Conference on Increasing Market and Planning Efficiency through Improved Software saw the release of 11 Optimal Power Flow and Formulation Papers by O'Neill *et al.* and discussions of how much of the $400 billion in energy sector revenues could be saved with improved optimization algorithms. O'Neill's estimate was $10 billion. This idea eventually led to the funding of ARPA-E's GO Competition and its several Challenges, each focused on an aspect of the problem more difficult than the last. The goal is to replace the approximate. linear (DC) optimization models currently used with accurate, physically correct non-linear (AC) models that can run as fast, or faster. A set of challenges was deemed the most effective way to stimulate interest and provide a fair comparison of the competing solvers. This requires a clear problem definition, a range of viable problem datasets, a consistent hardware platform, a pre-determined evaluation method to do the scoring, and a set of rules.

Challenge 1, from 2018 to 2019, focused on the basic Security Constrained AC Optimal Power Flow problem (SCOPF). The Challenge utilized sets of unique datasets generated by the ARPA-E GRID DATA program. Each dataset consisted of a collection of power system network models of different sizes with associated operating scenarios (snapshots in time defining instantaneous power demand, renewable generation, generator and line availability, etc.). Some datasets were synthetic and others were from industry. On July 24, 2018, ARPA-E issued a funding opportunity announcement (FOA, (DE-FOA-0001952) of up to $5 million for teams to participate, since some of the most talented organizations, e.g., US FFRDCs, could not compete without funding. Secretary of Energy Rick Perry announced the 18 FOA winners on October 31, 2018. Challenge 1 proceeded with 3 Trial Events (April 15, July 19, and Sept. 13) followed by a Final Event Oct. 31, 2019, which awarded $3.4 million dollars to 10 teams, announced by U.S. Energy Secretary Dan Brouillette on Feb. 12, 2020. The prize winners in Challenge 1 who were also FOA winners were required to spend the funds on Challenge 2.

Challenge 2, from 2020 to 2021, expanded upon the problem posed in Challenge 1 by adding adjustable transformer tap ratios, phase shifting transformers, switchable shunts, price-responsive demand, ramp rate constrained generators and loads, and fast-start unit commitment. Furthermore, Challenge 2 is a maximization problem while Challenge 1 was a minimization problem. Specifically, the economic surplus, defined as the benefit of serving load minus the cost of generation, is being maximized. It is expected that the objective value of a given solution should be positive, representing economic gain, but negative objectives from poor solutions are possible. Additionally, Divisions 3 and 4 within the competition permitted on/off switching of transmission lines (Divisions 1 and 2 did not). Challenge 2 awarded $2.4 million in prizes to 9 teams.

While Challenge 3 was being developed in 2022, the entrants were invited to see if they could produce better solutions to 84 synthetic Challenge 2 datasets with no restrictions on time, hardware, or algorithms. This Challenge 2: Monarch of the Mountain (CH2-MoM) awarded two teams $440,000. Improved results were found for all 84 datasets with improvements running from 6.7% to 0.00039%.

Challenge 3, from 2022 to 2023, expanded the Challenge 2 problem further by using multiperiod dynamic markets, including advisory models for extreme weather events, day-ahead markets, and the real-time markets with an extended look-ahead. These problems included active bid-in demand and topology optimization. Challenge 3 awarded $3.0 million in prized to 8 teams.

Elements

Providing a level playing field to determine the best teams, the Competition focused on several key elements. The basic evaluation process is to take an entrant's solver from GitHub, run it on PNNH hardware against a set of problem datasets (producing a solution for each), evaluate the solution file(s) generating scoring information, and return the results to the entrant.

Problem Formulation

A concise, consistent problem formulation is one of the key elements. These were large, mathematically intense documents that were reviewed (and corrected as necessary; corrections are listed in the Change Log appendix) by all the entrants. Several trial events were held to make sure that all entrants understood the problem and the evaluation procedure, which includes specifying the correct runtime environment.

Challenge 1: 4/9/2019 PDF 82 pages, 210 equations
Challenge 2: 5/31/2021 PDF 97 pages 299 equations
Challenge 3: 5/15/2023 PDF 62 pages 320 equations. The 1/22/2024 PDF, 67 pages 328 equations, which was revised at the conclusion of Challenge 3, includes an expanded Introduction and a Problem Description section.

Data Format—Input and Output

The dataset format, both input and output, was thoroughly documented and annotated to assure that all Entrants were on the same page. The first two Challenges used a mixture of commercial (Siemens PSS®E style) and Competition specific formats. Challenge 3 used only a Competition developed Json format. The datasets themselves were provided by collaborators from the ARPA-E GRID DATA project (Georgia Tech, NREL, PNNL, Texas A&M, and the University of Wisconsin-Madison). PNNL provided the industry datasets for Challenge 1 and Georgia Tech the industry datasets for Challenges 2 and 3. Proprietary CEII industry datasets could be used because the computation was behind a firewall under the complete control of PNNL, which masked the results returned to the entrants except for just the objective values, *i.e.*, only the solver behind the firewall could see the data and the results.

PNNL also developed dataset validation software that assured that the data from the dataset providers was valid and had feasible solutions. This software was made publicly available, and some entrants adapted it for reading the input. PNNL also developed software that reads the solution files and produces the objective score (and additional analysis information) and was likewise made available to be adapted for solution writing.

Evaluation Platform

To score an entrant's software, it was run on the same hardware and software environment as every other entrants' software. The software had dedicated access to the requested number of nodes, but inter-node communication shared the network fabric of the entire machine. The global file system was another shared component. Evaluation runs that completed well within the time limit showed very consistent results, but near the time limit there could be variations in the results by a few percent. When this variation could impact score ranking, multiple runs were made to seek a consensus result. For privacy, the Entrant was required to install a PNNL provided SSH public key on GitHub. The location of the GitHub repository was provided during the Team registration. Only a properly registered Team member may make a submission.

Website

In addition to providing information relevant to the Competition, including downloading documentation and publicly available datasets, the Website registered the entrants and provided the mechanism to initiate the scoring of a solver. There were two types of scoring submission: the Event submission where the entrant provided the solver and runtime information; and the Sandbox submission. An Event submission would run all the scenario datasets associated with an Event (including all Divisions) while the Sandbox submission would run only the single scenario and Division(s) requested. Results from the Event submission would go to the appropriate leaderboard, but the Sandbox submission results would return only to the entrant. The Sandbox submission would also return status updates during the run.

Sandbox

The Sandbox provided Entrants the opportunity to become familiar with the Competition platform. Submissions against available datasets are executed, results are evaluated, scored, and relevant data, which could include log files, solution files, and evaluation results for synthetic datasets, returned to the Entrant.

To run individual scenarios in the Sandbox, an Entrant must specify the appropriate dataset, model, scenario and runtime information (language, libraries, *etc*.) in a text box submission.conf on the submission page or in a submission.conf text file in the GitHub repository.

Input Parameters

For Challenge 1 and 2, an Entrants' code was executed twice, once to determine the base case solution and once to determine contingency solutions. Strict wall clock timing was determined by the TimeLimitInSeconds parameter in both cases. Two versions of the code were needed: Code1, which produced the file solution_BASECASE.txt; and Code2, which produced a file

solution_labelk.txt for each contingency k with labelk. The solution files were created at the same directory level that the run commands are executed.

For Challenge 3, only a single Code1 was needed to produce the file solution.json.
The naming, invocation procedure, and other language dependencies were described for each language supported. The language was determined in the submission.conf file on GitHub, which is specified during the submission process along with the submission title and notes.
Five input parameters were passed to Code1:

InFile1 = scenario_nnn.json (Problem description data)

TimeLimitInSeconds = the amount of wall-clock time in seconds before the execution will be terminated. The Challenge 3 value, in seconds, is either 600 (10 minutes) for Division 1, 7200 (120 minutes or 2 hours) for Division 2, and 14400 (240 minutes or 4 hours) for Division 3 when executing *Code1*. Divisions 4, 5, and 6, which use rank scoring, use the results from Divisions 1, 2, and 3, which use objective scoring.

Division = 1, 2, or 3. All Challenge 3 Divisions use Objective Function Scoring where the objective is defined by parameter "z" in the Formulation Document. The input files for a given scenario are different for each division.

NetworkModel = C3VvNxxxxxDn is a string identifying the Network Model of the input files where V is either S (Sandbox) or E (Event), v is 1-3 character string, N marks the beginning of a 5 digit number of buses in the network model, and D marks the beginning of a 1 digit Division number (1, 2, or 3). These are the first characters of the Dataset Network Model folder name.

AllowSwitching = Either a 1 (switching is allowed) or 0 (switching is NOT allowed). This is primarily for post-Challenge analysis to study the impact of switching. In order to ensure that the AllowSwitching argument is implemented and able to be used in post-Challenge analysis, there were some Event runs contributing to Challenge 3 prize awards using “AllowSwitching=1”, and there were some using “AllowSwitching=0”. A solution is said to use switching if the on/off status of some AC line or transformer in some time interval is different from the initial status specified in the problem data. If a code is called with “AllowSwitching=0”, and the solution evaluator determines that the solution does use switching, then the solution will be deemed infeasible.

submission.conf

The file, submission.conf (all lower case), must be added to the root the GitHub repository, but is overriden by the contents of the submission.conf text box provided at the time of submission. This text file determines which dataset is run and permits additional control of the runtime environment. All entries are optional and defaults are assumed if not specified. All entries are case sensitive. Dataset specific options are ignored for Event submissions; these will be

determined by the Event. The Challenge 3 parameters and their possible values (default underlined, options separated by |) are:

dataset=[S0.1 | S1.1 | S2.1 | S3.1 | S4X | E1.1 | E2.1 | E3.1]

model=Network_Model_Name, where Network_Model_Name (no spaces around equal sign) is one of the network model names belonging to a given submission. For example, one of the network models in the Sandbox submission set (C3_Sandbox) C3SvNxxxxxDn where C3 indicates this is a Challenge 3 dataset, Sv indicates it is a Sandbox dataset from release v, Nxxxxx indicates that the Network has xxxxx buses, and Dn indicates the Division number (1, 2, or 3 for Challenge 3; no Dn for S0). The purpose of this field is solely for identification of the problem being solved. The model parameter is required EXCEPT during an Event when the choice is made by the platform. Example: model=C3S0.1N00014

scenario=scenario_number, where scenario_number (no spaces around equal sign) is a valid integer number of a scenario in the dataset selected by the model parameter above. The number of scenarios for available for a given model is given on the dataset page for the Sandbox datasets. For Events, all scenarios will be run for you. For the model example above, "model=C3S0.1N00014", the choice is scenario 3, i.e., "scenario=3". The scenario parameter is required EXCEPT during an Event when the choice is made by the platform.

language = [cpp | EXE | GAMS | Java | Julia | Python]

experiment = [SW0 | SW1], where SW0 indicates that the **AllowSwitching** parameter above is set to 0 (switching not allowed and if detected will result in an infeasible determination for the solution). The default is SW1, i.e., switching is allowed.

The default time limits may be overridden up to the maximum value of 24 hours for Sandbox runs. This does not apply to Event runs. Time is always specified in seconds.

div1_timelimit_sec = [600 | value <86,400 (24 hours)]

div2_timelimit_sec = [7,200 | value <86,400 (24 hours)]

div3_timelimit_sec = [14,400 | value <86,400 (24 hours)]

Choose specific versions of software (language, solver, MPI library) either through modules variable as below or through exporting environment variables in consultation with GO Operations Team.

modules=<space separated modules>

Here are a few commonly used modules, but other modules, versions and library packages are also possible: python/2.7.14; python/3.7.0; gcc/7.3.0; cmake/3.15.3; intel/18.0.0; intelmpi/2019u4; gcc/4.8.5 openmpi/2.1.1. There are other modules for Gurobi, Julia/JumP, Ipopt, MPI (various flavors: openmpi/2.1.1; openmpi/3.1.3; openmpi/4.0.2; openmpi/4.0.6; openmpi/4.1.0; openmpi/4.1.1; openmpi/4.1.3; openmpi/4.1.4; gcc/10.2.0; gcc/10.3.0;

gcc/11.2.0; gcc/4.4.7; gcc/4.8.5; gcc/5.2.0; gcc/7.1.0; gcc/7.3.0; gcc/7.5.0; gcc/8.1.0; gcc/8.4.0; gcc/9.1.0; mvapich2/2.3a; mvapich2/2.3.2; mvapich2/2.3.5; mvapich2/2.3.5.fthread; mvapich2/2.3.6; mvapich2/2.3.6.fthread; mvapich2/2.3.7), Java/Scala, and GAMS.

The list of supported runtime environments is extensive, and no Entrant was ever turned away because their solver could not be supported. Getting a correct set of submission.conf parameters is nontrivial and no Entrant got it right the first time, but, with help from the GO Operations Team, all teams were eventually able to make successful submissions. Further practice using the Sandbox correlated with success by the winning teams.

Event 1 Experience

The initial Sandbox datasets (S0) were released 12/18/2022, a little over a month before the Event 1 submission window of January 25-27, 2023, and consisted of 4 small network models of 3-, 14-, 37-, and 73-buses with one scenario for each Division (12 total). These scenarios were later updated on 8/4/2023 (S0.1) to comply with the positivity requirement of the Problem Formulation introduced 5/15/2023.

A few days after S0 was released, the S1 datasets were released on 12/22/2022. This dataset contained 4 larger network models consisting of 600-, 1576-, 4200- and 6049-buses, again, each with a single scenario for each Division (12 scenarios total). As with S0, these scenarios were later updated on 8/7/2023 (S1.1) to comply with the positivity requirement (no new scenarios).

At the start of Event 1 there were 24 synthetic and no industry scenarios to practice on.

As shown in Table 1, the teams that practiced more (made more Sandbox submissions), tended to do better. The Total Score Rank is based on summing all the scores (objective values) from running the 281 scenarios in Event 1. The Best Score Rank counts the number of scenarios where the team got the best score. See the Event 1 Leaderboard for details. For the 12 teams (this includes the Benchmark) evaluated in Event 1, this was a minimum of 3,372 evaluations (some scenarios were run more than once for a given team).

Table 1. C3 Event 1 Submissions

Team name	Sandbox submissions	Total Score Rank	Best Score Rank
TIM-GO	268	3	2
Argonauts	135	6	5
Electric-Stampede	120	2	4
Gatorgar	53	5	no firsts
LLGoMax	47	no score	no firsts
GOT-BSI-OPF	44	1	3
Artelys_Columbia	23	4	1
PACE	23	no score	no firsts
PGWOpt	21	no score	no firsts
YongOptimization	16	no score	no firsts
The Blackouts	15	no score	no firsts

Total Submissions	**765**

Event 2 Experience

The 257 Event 1 synthetic scenarios were released 2/14/2023, about two weeks after Event 1 began. All scenarios complied with the positivity requirement. On 3/16/2023, the 6 S2 (positivity requirement compliant) datasets using the 2000- and 6717-bus networks were released ahead of the Event 2 submission period of April 13-14, 2023.

At the start of Event 2 there were 24+257+6 (287) synthetic scenarios to practice on.

Event 2 proceeded with a total of 197 scenarios (167 synthetic). This was further broken down into 27 unique Industry datasets and 148 unique synthetic datasets; 3 industry and 19 synthetic scenarios were run with and without switching.

As in Table 1, Table 2 shows that teams that made more Sandbox submissions tended to do better. There were 197 scenarios used in Event 2, leading to a minimum of 2,561 runs.

Table 2. C3 Event 2 Submissions

Team name	Sandbox submissions	Total Score Rank	Best Score Rank
Electric-Stampede	1179	1	3
GravityX	138	no E2 sub.	no E2 sub.
YongOptimization	109	5	2
LLGoMax	58	8	no firsts
Artelys_Columbia	55	4	4
TIM-GO	53	2	1
GOT-BSI-OPF	46	3	5 (tie)
PGWOpt	31	no score	no firsts
quasiGrad	28	no score	no firsts
Gatorgar	22	7	no firsts
Argonauts	13	9	no firsts
PACE	5	no score	no firsts
The Blackouts	5	6	5 (tie)
Total Submissions	**1742**		

Event 3 Experience

The 148 unique synthetic Event 2 scenarios were released 5/10/2023 about a month ahead of the June 15-16, 2023, Event 3 submission window. All the datasets satisfied the positivity requirement except the 2000- and 6717-bus networks, which were released on 5/15/2023 as E2.1. The 6708-bus industry dataset with 27 scenarios was not released but was available to run in the Sandbox. On 6/6/2023, 41 S3.1 (positivity requirement compliant) datasets using the 14-, 37-,

1576-, 2000-, 8316, and 23643-bus networks were released, but the next day it was discovered that 3 of the scenarios had invalid data, leaving 37 valid S3.1 scenarios.

At the start of Event 3 there were 287+148+37 (472) synthetic scenarios and 27 industry scenarios to practice on.

Event 3 proceeded with a total of 139 scenarios (130 synthetic). This was further broken down into 9 unique industry datasets and 100 unique synthetic datasets; 0 industry and 30 synthetic scenarios were run with and without switching. Prizes were awarded for best cumulative score in each Division, but not for the number of best scores.

As in Events 1 and 2, Table 3 shows that teams that made more Sandbox submissions tended to do better. There were 137 scenarios used in Event 3, with the large 23643-bus case being run with standard and 72-hour time limits, leading to a minimum of 1,960 runs. For Event 3 the number of Sandbox runs exceeds the number of Event runs.
It is worthwhile commenting on why The Blackouts team scored better than the number of Sandbox runs would indicate. This team had ample local compute resources and made just enough Sandbox runs to ensure the runs would proceed correctly. Their strategy was to use the full amount of allotted time for each run to assure the best result possible. Unfortunately, that meant they encountered many timeout situations with no solution file being produced. On their local system, the shutdown mechanism operated differently than the Evaluation Platform, producing more solution files locally than in the Competition. This was not discovered until the Event 4 data was analyzed.

Table 3. C3 Event 3 Submissions

Team name	Sandbox submissions	$k prize	Best Score Rank
Electric-Stampede	1926	115	6 (tie)
TIM-GO	1013	75	3 (tie)
YongOptimization	463	95	2
GravityX	415	120	3 (tie)
GOT-BSI-OPF	257	125	1
Artelys_Columbia	127	70	6 (tie)
quasiGrad	63		no firsts
Gatorgar	56		no firsts
LLGoMax	39		no firsts
The Blackouts	33		5
PACE	8		no firsts
PGWOpt	8		no firsts
Argonauts	4		no firsts
total	**4412**	**600**	

Event 4 Experience

The 98 unique synthetic Event 3 scenarios were released 6/29/2023 about 2 months ahead of the August 31 to September 4, 2023, Event 4 submission window. All the E3 datasets satisfied the positivity requirement.

On 8/3/2023, 9 S4X (switching enhanced) datasets using the 617-bus Division 1 network were released and on 8/9/2023 another 4 S4X datasets using the 73-bus Division 2 network was released.

On 8/21/2023, 16 new S4 (called S3.1 with the new date and positivity requirement compliant) datasets using the 73-, 1576-, 2000-, 6049-, 6717-, and 8316-bus networks were released.

At the start of Event 4 there were 472+98+9+4+16 (599) synthetic scenarios and 36 industry scenarios to practice on.

Event 4 proceeded with a total of 669 scenarios (626 synthetic). This was further broken down into 39 unique industry datasets and 591 unique synthetic datasets; 4 industry and 35 synthetic scenarios were run with and without switching. Prizes were awarded for best cumulative score and number of best scores in each Division.

Table 4 again shows that teams that made more Sandbox submissions tended to do better. The exception for The Blackouts has already been noted. The exception for Electric-Stampede just shows that it takes more than practice. They maintain they were not building machine learning training information.

There were 669 scenarios used in Event 3, with the large 23643-bus case being run with just the 72-hour time this time, leading to a minimum of 17,928 runs. The total number of Sandbox runs for the entire Challenge 3 (16,876) is of the same order as the number of Event 4 runs.

Table 4. C3 Event 4 Submissions

Team name	Sandbox submissions	$k prize
Electric-Stampede	2957	
TIM-GO	2637	520
YongOptimization	2052	550
GravityX	551	320
Artelys_Columbia	431	320
GOT-BSI-OPF	313	360
Occam's razor	286	130
quasiGrad	205	
PGWOpt	173	
Quasimodo	132	
LLGoMax	94	
Gatorgar	84	

The Blackouts		35	200
PACE		7	
	total	**9957**	**2400**

Post Event 4

The 591 unique synthetic Event 5 scenarios were released as a set of 19 zip files occupying 3.6 GB on October 2, 2023, and Sandbox submissions were turned back on. Since then, 3 teams have made additional Sandbox submissions.

YongOptimization made 32 submissions with the last on November 14, 2023.

Electric-Stampede made 123 submissions with the last on December 11, 2023.

TIM-GO made 53 submissions with the last on March 19, 2024.