angr Documentation

README

angr is a multi-architecture binary analysis toolkit, with the capability to perform dynamic symbolic execution (like Mayhem, KLEE, etc.) and various static analyses on binaries. If you'd like to learn how to use it, you're in the right place!

We've tried to make using angr as pain-free as possible - our goal is to create a user-friendly binary analysis suite, allowing a user to simply start up iPython and easily perform intensive binary analyses with a couple of commands. That being said, binary analysis is complex, which makes angr complex. This documentation is an attempt to help out with that, providing narrative explanation and exploration of angr and its design.

Several challenges must be overcome to programmatically analyze a binary. They are, roughly:

- Loading a binary into the analysis program.
- Translating a binary into an intermediate representation (IR).
- Performing the actual analysis. This could be:
  - A partial or full-program static analysis (i.e., dependency analysis, program slicing).
  - A symbolic exploration of the program's state space (i.e., "Can we execute it until we find an overflow?").
  - Some combination of the above (i.e., "Let's execute only program slices that lead to a memory write, to find an overflow.")

angr has components that meet all of these challenges. This book will explain how each one works, and how they can all be used to accomplish your evil goals.

Get Started

Installation instructions can be found here.

To dive right into angr's capabilities, start with the top level methods and read forward from there.

A searchable HTML version of this documentation is hosted at docs.angr.io, and an HTML API reference can be found at angr.io/api-doc.

If you enjoy playing CTFs and would like to learn angr in a similar fashion, angr_ctf will be a fun way for you to get familiar with much of the symbolic execution capability of angr. The angr_ctf repo is maintained by @jakespringer.

Citing angr

If you use angr in an academic work, please cite the papers for which it was developed:

```
@article{shoshitaishvili2016state,
  title={SoK: (State of) The Art of War: Offensive Techniques in Binary Analysis},
  author={Shoshitaishvili, Yan and Wang, Ruoyu and Salls, Christopher and Stephens, Nick a
  booktitle={IEEE Symposium on Security and Privacy},
  year={2016}
}

@article{stephens2016driller,
  title={Driller: Augmenting Fuzzing Through Selective Symbolic Execution},
  author={Stephens, Nick and Grosen, John and Salls, Christopher and Dutcher, Audrey and W
  booktitle={NDSS},
  year={2016}
}

@article{shoshitaishvili2015firmalice,
  title={Firmalice - Automatic Detection of Authentication Bypass Vulnerabilities in Binar
  author={Shoshitaishvili, Yan and Wang, Ruoyu and Hauser, Christophe and Kruegel, Christo
  booktitle={NDSS},
  year={2015}
}
```

Support

To get help with angr, you can ask via:

- the slack channel: angr.slack.com, for which you can get an account here.
- opening an issue on the appropriate github repository
- the mailing list: angr@lists.cs.ucsb.edu

Going further:

You can read this paper, explaining some of the internals, algorithms, and used techniques to get a better understanding on what's going on under the hood.

Introductory Errata

Installing

Installing angr

angr is a python library, so it must be installed into your python environment before it can be used. It is built for Python 3: Python 2 support is not feasible due to the looming EOL and the small size of our team.

We highly recommend using a python virtual environment to install and use angr. Several of angr's dependencies (z3, pyvex) require libraries of native code that are forked from their originals, and if you already have libz3 or libVEX installed, you definitely don't want to overwrite the official shared objects with ours. In general, don't expect support for problems arising from installing angr outside of a virtualenv.

Dependencies

All of the python dependencies should be handled by pip and/or the setup.py scripts. You will, however, need to build some C to get from here to the end, so you'll need a good build environment as well as the python development headers. At some point in the dependency install process, you'll install the python library cffi, but (on linux, at least) it won't run unless you install your operating system's libffi package.

On Ubuntu, you will want: `sudo apt-get install python3-dev libffi-dev build-essential virtualenvwrapper`. If you are trying out angr Management, you will also need the PySide 2 requirements.

Most Operating systems, all *nix systems

`mkvirtualenv --python=$(which python3) angr && pip install angr` should usually be sufficient to install angr in most cases, since angr is published on the Python Package Index.

Fish (shell) users can either use virtualfish or the virtualenv package: `vf new angr && vf activate angr && pip install angr`

Failing that, you can install angr by installing the following repositories, in order, from https://github.com/angr:

- archinfo
- pyvex
- claripy
- cle
- angr

Mac OS X

`pip install angr` should work, but there are some caveats.

angr requires the `unicorn` library, which (as of this writing) `pip` must build from source on macOS, even though binary distributions ("wheels") exist on other platforms. Building `unicorn` from source requires Python 2, so will fail inside a virtualenv where `python` gets you Python 3. If you encounter errors with `pip install angr`, you may need to first install `unicorn` separately, pointing it to your Python 2:

```
UNICORN_QEMU_FLAGS="--python=/path/to/python2" pip install unicorn  # Python 2 is probably
```

Then retry `pip install angr`.

If this still doesn't work and you run into a broken build script with Clang, try using GCC.

```
brew install gcc
CC=/usr/local/bin/gcc-8 UNICORN_QEMU_FLAGS="--python=/path/to/python2" pip install unicorn
pip install angr
```

After installing angr, you will need to fix some shared library paths for the angr native libraries. Activate your virtual env and execute the following lines. A script is provided in the angr-dev repo.

```
PYVEX=`python3 -c 'import pyvex; print(pyvex.__path__[0])'`
UNICORN=`python3 -c 'import unicorn; print(unicorn.__path__[0])'`
ANGR=`python3 -c 'import angr; print(angr.__path__[0])'`

install_name_tool -change libunicorn.1.dylib "$UNICORN"/lib/libunicorn.dylib "$ANGR"/lib/a
install_name_tool -change libpyvex.dylib "$PYVEX"/lib/libpyvex.dylib "$ANGR"/lib/angr_nati
```

Windows

As usual, a virtualenv is very strongly recommended. You can use either the virtualenv-win or virtualenv packages for this.

angr can be installed from pip on Windows, same as above: `pip install angr`. You should not be required to build any C code with this setup, since wheels (binary distributions) should be automatically pulled down for angr and its dependencies.

Nix/NixOS

angr is available via the Nix package manager and on NixOS, using the Nix User Repository.

First, make NUR available to your user:

```
cat << __EOF__ > ~/.config/nixpkgs/config.nix
{
  packageOverrides = pkgs: {
    nur = import (builtins.fetchTarball "https://github.com/nix-community/NUR/archive/maste
      inherit pkgs;
    };
  };
}
__EOF__
```

Then, to obtain a nix-shell with the `angr` Python package:

```
nix-shell -p 'python3.withPackages(ps: with ps; [ nur.repos.angr.python3Packages.angr ])'
```

More information on angr/nixpkgs.

Development install

There is a special repository `angr-dev` with scripts to make life easier for angr developers. You can set up angr in development mode by running:

```
git clone https://github.com/angr/angr-dev
cd angr-dev
./setup.sh -i -e angr
```

This creates a virtualenv (`-e angr`), checks for any dependencies you might need (`-i`), clones all of the repositories and installs them in editable mode. `setup.sh` can even create a PyPy virtualenv for you (replace `-e` with `-p`), resulting in significantly faster performance and lower memory usage.

You can branch/edit/recompile the various modules in-place, and it will automatically reflect in your virtual environment.

Development install on windows

The angr-dev repository has a setup.bat script that creates the same setup as above, though it's not as magical as setup.sh. Since we'll be building C code, you must be in the visual studio developer command prompt. *Make sure that if you're using a 64-bit python interpreter, you're also using the 64-bit build tools* (`VsDevCmd.bat -arch=x64`)

```
pip install virtualenv
git clone https://github.com/angr/angr-dev
cd angr-dev
virtualenv -p "C:\Path\To\python3\python.exe" env
env\Scripts\activate
setup.bat
```

You may also substitute the use of `virtualenv` above with the `virtualenvwrapper-win` package for a more streamlined experience.

Docker install

For convenience, we ship a Docker image that is 99% guaranteed to work. You can install via docker by doing:

```
# install docker
curl -sSL https://get.docker.com/ | sudo sh

```

```
# pull the docker image
sudo docker pull angr/angr

# run it
sudo docker run -it angr/angr
```

Synchronization of files in and out of docker is left as an exercise to the user (hint: check out `docker run -v`).

Modifying the angr container

You might find yourself needing to install additional packages via apt. The vanilla version of the container does not have the sudo package installed, which means the default user in the container cannot escalate privilege to install additional packages.

To over come this hurdle, use the following docker command to grant yourself root access:

```
# assuming the docker container is running
# with the name "angr" and the instance is
# running in the background.
docker exec -ti -u root angr bash
```

Troubleshooting

libgomp.so.1: version GOMP_4.0 not found, or other z3 issues

This specific error represents an incompatibility between the pre-compiled version of libz3.so and the installed version of `libgomp`. A Z3 recompile is required. You can do this by executing:

```
pip install -I --no-binary z3-solver z3-solver
```

No such file or directory: 'pyvex_c'

Are you running Ubuntu 12.04? If so, please stop using a 6 year old operating system! Upgrading is free!

You can also try upgrading pip (`python -m pip install -U pip`), which might solve the issue.

AttributeError: 'FFI' object has no attribute 'unpack'

You have an outdated version of the `cffi` Python module. angr now requires at least version 1.7 of cffi. Try `pip install --upgrade cffi`. If the problem persists, make sure your operating system hasn't pre-installed an old version of cffi, which pip may refuse to uninstall. If you're using a Python virtual environment with the pypy interpreter, ensure you have a recent version of pypy, as it includes a version of cffi which pip will not upgrade.

angr has no attribute Project, or similar

If you can import angr but it doesn't seem to be the actual angr module... did you accidentally name your script `angr.py`? You can't do that. Python does not work that way.

AttributeError: 'module' object has no attribute 'KS_ARCH_X86'

You have the `keystone` package installed, which conflicts with the `keystone-engine` package (an optional dependency of angr). Please uninstall `keystone`. If you would like to install `keystone-engine`, please do it with `pip install --no-binary keystone-engine keystone-engine`, as the current pip distribution is broken.

No such file or directory: 'libunicorn.dylib'

(alternate error message: `Cannot use 'python', Python 2.4 or later is required. Note that Python 3 or later is not yet supported.`)

You need to define the `UNICORN_QEMU_FLAGS` environment variable for `pip`. See the section above on installing for macOS.

pthread check failed: Make sure to have the pthread libs and headers installed.

(macOS) Try using GCC instead of Clang; see the section above on installing for macOS.

How to Contribute

Reporting Bugs

If you've found something that angr isn't able to solve and appears to be a bug, please let us know!

1. Create a fork off of angr/binaries and angr/angr
2. Give us a pull request with angr/binaries, with the binaries in question
3. Give us a pull request for angr/angr, with testcases that trigger the binaries in `angr/tests/broken_x.py`, `angr/tests/broken_y.py`, etc

Please try to follow the testcase format that we have (so the code is in a test_blah function), that way we can very easily merge that and make the scripts run.
An example is:

```python
def test_some_broken_feature():
    p = angr.Project("some_binary")
    result = p.analyses.SomethingThatDoesNotWork()
    assert result == "what it should *actually* be if it worked"

if __name__ == '__main__':
    test_some_broken_feature()
```

This will *greatly* help us recreate your bug and fix it faster.
The ideal situation is that, when the bug is fixed, your testcases passes (i.e., the assert at the end does not raise an AssertionError).
Then, we can just fix the bug and rename `broken_x.py` to `test_x.py` and the testcase will run in our internal CI at every push, ensuring that we do not break this feature again.

Developing angr

These are some guidelines so that we can keep the codebase in good shape!

Coding style

We try to get as close as the PEP8 code convention as is reasonable without being dumb. If you use Vim, the python-mode plugin does all you need. You can also manually configure vim to adopt this behavior.

Most importantly, please consider the following when writing code as part of angr:

- Try to use attribute access (see the `@property` decorator) instead of getters and setters wherever you can. This isn't Java, and attributes enable tab completion in iPython. That being said, be reasonable: attributes should be fast. A rule of thumb is that if something could require a constraint solve, it should not be an attribute.
- Use our `.pylintrc` from the angr-dev repo. It's fairly permissive, but our CI server will fail your builds if pylint complains under those settings.
- DO NOT, under ANY circumstances, `raise Exception` or `assert False`. **Use the right exception type**. If there isn't a correct exception type, subclass the core exception of the module that you're working in (i.e., `AngrError` in angr, `SimError` in SimuVEX, etc) and raise that. We catch, and properly handle, the right types of errors in the right places, but `AssertionError` and `Exception` are not handled anywhere and force-terminate analyses.
- Avoid tabs; use space indentation instead. Even though it's wrong, the de facto standard is 4 spaces. It is a good idea to adopt this from the beginning, as merging code that mixes both tab and space indentation is awful.
- Avoid super long lines. It's okay to have longer lines, but keep in mind that long lines are harder to read and should be avoided. Let's try to stick to **120 characters**.
- Avoid extremely long functions, it is often better to break them up into smaller functions.
- Always use `_` instead of `__` for private members (so that we can access them when debugging). *You* might not think that anyone has a need to call a given function, but trust us, you're wrong.

Documentation

Document your code. Every *class definition* and *public function definition* should have some description of:

- What it does.

- What are the type and the meaning of the parameters.
- What it returns.

Class docstrings will be enforced by our linter. Do *not* under any circumstances write a docstring which doesn't provide more information than the name of the class. What you should try to write is a description of the environment that the class should be used in. If the class should not be instantiated by end-users, write a description of where it will be generated and how instances can be acquired. If the class should be instanciated by end-users, explain what kind of object it represents at its core, what behavior is expected of its parameters, and how to safely manage objects of its type.

We use Sphinx to generate the API documentation. Sphinx supports docstrings written in ReStructured Text with special keywords to document function and class parameters, return values, return types, members, etc.

Here is an example of function documentation. Ideally the parameter descriptions should be aligned vertically to make the docstrings as readable as possible.

```python
def prune(self, filter_func=None, from_stash=None, to_stash=None):
    """
    Prune unsatisfiable paths from a stash.

    :param filter_func: Only prune paths that match this filter.
    :param from_stash:  Prune paths from this stash. (default: 'active')
    :param to_stash:    Put pruned paths in this stash. (default: 'pruned')
    :returns:           The resulting PathGroup.
    :rtype:             PathGroup
    """
```

This format has the advantage that the function parameters are clearly identified in the generated documentation. However, it can make the documentation repetitive, in some cases a textual description can be more readable. Pick the format you feel is more appropriate for the functions or classes you are documenting.

```python
def read_bytes(self, addr, n):
    """
    Read `n` bytes at address `addr` in memory and return an array of bytes.
    """
```

Unit tests

If you're pushing a new feature and it is not accompanied by a test case it **will be broken** in very short order. Please write test cases for your stuff.

We have an internal CI server to run tests to check functionality and regression on each commit. In order to have our server run your tests, write your tests in a format acceptable to nosetests in a file matching `test_*.py` in the `tests` folder of the appropriate repository. A test file can contain any number of functions of the form `def test_*():` or classes of the form `class Test*(unittest.TestCase):`. Each of them will be run as a test, and if they raise any exceptions or assertions, the test fails. Do not use the `nose.tools.assert_*` functions, as we are presently trying to

migrate to `nose2`. Use `assert` statements with descriptive messages or the `unittest.TestCase` assert methods.
Look at the existing tests for examples. Many of them use an alternate format where the `test_*` function is actually a generator that yields tuples of functions to call and their arguments, for easy parametrization of tests.

Finally, do not add docstrings to your test functions.

What to Contribute

angr is a huge project, and it's hard to keep up. Here, we list some big TODO items that we would love community contributions for in the hope that it can direct community involvement. They (will) have a wide range of complexity, and there should be something for all skill levels!

We tag issues on our github repositories that would be good for community involvement as "Help wanted". To see the exhaustive list of these, use this github search!

Documentation

There are many parts of angr that suffer from little or no documentation. We desperately need community help in this area.

API

We are always behind on documentation. We've created several tracking issues on github to understand what's still missing:

1. angr
2. claripy
3. cle
4. pyvex

GitBook

This book is missing some core areas. Specifically, the following could be improved:

1. Finish some of the TODOs floating around the book.
2. Organize the Examples page in some way that makes sense. Right now, most of the examples are very redundant. It might be cool to have a simple table of most of them so that the page is not so overwhelming.

angr course

Developing a "course" of sorts to get people started with angr would be really beneficial. Steps have already been made in this direction here, but more expansion would be beneficial.

Ideally, the course would have a hands-on component, of increasing difficulty, that would require people to use more and more of angr's capabilities.

Research re-implementation

Unfortunately, not everyone bases their research on angr ;-). Until that's remedied, we'll need to periodically implement related work, on top of angr, to make it reusable within the scope of the framework. This section lists some of this related work that's ripe for reimplementation in angr.

Redundant State Detection for Dynamic Symbolic Execution

Bugrara, et al. describe a method to identify and trim redundant states, increasing the speed of symbolic execution by up to 50 times and coverage by 4%. This would be great to have in angr, as an ExplorationTechnique. The paper is here: http://nsl.cs.columbia.edu/projects/minestrone/papers/atc13-bugrara.pdf

In-Vivo Multi-Path Analysis of Software Systems

Rather than developing symbolic summaries for every system call, we can use a technique proposed by S2E for concretizing necessary data and dispatching them to the OS itself. This would make angr applicable to a *much* larger set of binaries than it can currently analyze.

While this would be most useful for system calls, once it is implemented, it could be trivially applied to any location of code (i.e., library functions). By carefully choosing which library functions are handled like this, we can greatly increase angr's scalability.

Development

We have several projects in mind that primarily require development effort.

angr-management

The angr GUI, angr-management needs a *lot* of work. Here is a non-exhaustive list of what is currently missing in angr-management:

- A navigator toolbar showing content in a program’s memory space, just like IDA Pro’s navigator toolbar.
- A text-based disassembly view of the program.
- Better view showing details in program states during path exploration, including modifiable register view, memory view, file descriptor view, etc.
- A GUI for cross referencing.

Exposing angr's capabilities in a usable way, graphically, would be really useful!

IDA Plugins

Much of angr's functionality could be exposed via IDA. For example, angr's data dependence graph could be exposed in IDA through annotations, or obfuscated values can be resolved using symbolic execution.

Additional architectures

More architecture support would make angr all the more useful. Supporting a new architecture with angr would involve:

1. Adding the architecture information to archinfo
2. Adding an IR translation. This may be either an extension to PyVEX, producing IRSBs, or another IR entirely.
3. If your IR is not VEX, add a `SimEngine` to support it.
4. Adding a calling convention (`angr.SimCC`) to support SimProcedures (including system calls)
5. Adding or modifying an `angr.SimOS` to support initialization activities.
6. Creating a CLE backend to load binaries, or extending the CLE ELF backend to know about the new architecture if the binary format is ELF.

ideas for new architectures:

- PIC, AVR, other embedded architectures
- SPARC (there is some preliminary libVEX support for SPARC here)

ideas for new IRs:

- LLVM IR (with this, we can extend angr from just a Binary Analysis Framework to a Program Analysis Framework and expand its capabilities in other ways!)
- SOOT (there is no reason that angr can't analyze Java code, although doing so would require some extensions to our memory model)

Environment support

We use the concept of "function summaries" in angr to model the environment of operating systems (i.e., the effects of their system calls) and library functions. Extending this would be greatly helpful in increasing angr's utility. These function summaries can be found here.

A specific subset of this is system calls. Even more than library function SimProcedures (without which angr can always execute the actual function), we have very few workarounds for missing system calls. Every implemented system call extends the set of binaries that angr can handle.

Design Problems

There are some outstanding design challenges regarding the integration of additional functionalities into angr.

Type annotation and type information usage

angr has fledgling support for types, in the sense that it can parse them out of header files. However, those types are not well exposed to do anything useful with. Improving this support would make it possible to, for example, annotate certain memory regions with certain type information and interact with them intelligently. Consider, for example, interacting with a linked list like this: `print state.mem[state.regs.rax].llist.next.next.value`.

(editor's note: you can actually already do this)

Research Challenges

Historically, angr has progressed in the course of research into novel areas of program analysis. Here, we list several self-contained research projects that can be tackled.

Semantic function identification/diffing

Current function diffing techniques (TODO: some examples) have drawbacks. For the CGC, we created a semantic-based binary identification engine (https://github.com/angr/identifier) that can identify functions based on testcases. There are two areas of improvement, each of which is its own research project:

1. Currently, the testcases used by this component are human-generated. However, symbolic execution can be used to automatically generate testcases that can be used to recognize instances of a given function in other binaries.
2. By creating testcases that achieve a "high-enough" code coverage of a given function, we can detect changes in functionality by applying the set of testcases to another implementation of the same function and analyzing changes in code coverage. This can then be used as a sematic function diff.

Applying AFL's path selection criteria to symbolic execution

AFL does an excellent job in identifying "unique" paths during fuzzing by tracking the control flow transitions taken by every path. This same metric can be applied to symbolic exploration, and would probably do a depressingly good job, considering how simple it is.

Overarching Research Directions

There are areas of program analysis that are not well explored. We list general directions of research here, but readers should keep in mind that these directions likely describe potential undertakings of entire PhD dissertations.

Process interactions

Almost all work in the field of binary analysis deals with single binaries, but this is often unrealistic in the real world. For example, the type of input that can be passed to a CGI program depend on pre-processing by a web server. Currently, there is no way to support the analysis of multiple concurrent processes in angr, and many open questions in the field (i.e., how to model concurrent actions).

Intra-process concurrency

Similar to the modeling of interactions between processes, little work has been done in understanding the interaction of concurrent threads in the same process. Currently, angr has no way to reason about this, and it is unclear from the theoretical perspective how to approach this.

A subset of this problem is the analysis of signal handlers (or hardware interrupts). Each signal handler can be modeled as a thread that can be executed at any time that a signal can be triggered. Understanding when it is meaningful to analyze these handlers is an open problem. One system that does reason about the effect of interrupts is FIE.

Path explosion

Many approaches (such as Veritesting) attempt to mitigate the path explosion problem in symbolic execution. However, despite these efforts, path explosion is still *the* main problem preventing symbolic execution from being mainstream.

angr provides an excellent base to implement new techniques to control path explosion. Most approaches can be easily implemented as Exploration Techniques and quickly evaluated (for example, on the CGC dataset).

Frequently Asked Questions

This is a collection of commonly-asked "how do I do X?" questions and other general questions about angr, for those too lazy to read this whole document.

If your question is of the form "how do I fix X issue", see also the Troubleshooting section of the install instructions.

Why is it named angr?

The core of angr's analysis is on VEX IR, and when something is vexing, it makes you angry.

How should "angr" be stylized?

All lowercase, even at the beginning of sentences. It's an anti-proper noun.

How can I get diagnostic information about what angr is doing?

angr uses the standard `logging` module for logging, with every package and submodule creating a new logger.

The simplest way to get debug output is the following:

```python
import logging
logging.getLogger('angr').setLevel('DEBUG')
```

You may want to use `INFO` or whatever else instead. By default, angr will enable logging at the `WARNING` level.

Each angr module has its own logger string, usually all the python modules above it in the hierarchy, plus itself, joined with dots. For example, `angr.analyses.cfg`. Because of the way the python logging module works, you can set the verbosity for all submodules in a module by setting a verbosity level for the parent module. For example, `logging.getLogger('angr.analyses').setLevel('INFO')` will make the CFG, as well as all other analyses, log at the INFO level.

Why is angr so slow?

It's complicated!

How do I find bugs using angr?

It's complicated! The easiest way to do this is to define a "bug condition", for example, "the instruction pointer has become a symbolic variable", and run symbolic exploration until you find a state matching that condition, then dump the input as a testcase. However, you will quickly run into the state explosion problem. How you address this is up to you. Your solution may be as simple as adding an `avoid` condition or as complicated as implementing CMU's MAYHEM system as an Exploration Technique.

Why did you choose VEX instead of another IR (such as LLVM, REIL, BAP, etc)?

We had two design goals in angr that influenced this choice:

1. angr needed to be able to analyze binaries from multiple architectures. This mandated the use of an IR

to preserve our sanity, and required the IR to support many architectures.
2. We wanted to implement a binary analysis engine, not a binary lifter. Many projects start and end with the implementation of a lifter, which is a time consuming process. We needed to take something that existed and already supported the lifting of multiple architectures.

Searching around the internet, the major choices were:

- LLVM is an obvious first candidate, but lifting binary code to LLVM cleanly is a pain. The two solutions are either lifting to LLVM through QEMU, which is hackish (and the only implementation of it seems very tightly integrated into S2E), or McSema, which only supported x86 at the time but has since gone through a rewrite and gotten support for x86-64 and aarch64.
- TCG is QEMU's IR, but extracting it seems very daunting as well and documentation is very scarce.
- REIL seems promising, but there is no standard reference implementation that supports all the architectures that we wanted. It seems like a nice academic work, but to use it, we would have to implement our own lifters, which we wanted to avoid.
- BAP was another possibility. When we started work on angr, BAP only supported lifting x86 code, and up-to-date versions of BAP were only available to academic collaborators of the BAP authors. These were two deal-breakers. BAP has since become open, but it still only supports x86_64, x86, and ARM.
- VEX was the only choice that offered an open library and support for many architectures. As a bonus, it is very well documented and designed specifically for program analysis, making it very easy to use in angr.

While angr uses VEX now, there's no fundamental reason that multiple IRs cannot be used. There are two parts of angr, outside of the `angr.engines.vex` package, that are VEX-specific:

- the jump labels (i.e., the `Ijk_Ret` for returns, `Ijk_Call` for calls, and so forth) are VEX enums.
- VEX treats registers as a memory space, and so does angr. While we provide accesses to `state.regs.rax` and friends, on the backend, this does `state.registers.load(8, 8)`, where the first `8` is a VEX-defined offset for `rax` to the register file.

To support multiple IRs, we'll either want to abstract these things or translate their labels to VEX analogues.

Why are some ARM addresses off-by-one?

In order to encode THUMB-ness of an ARM code address, we set the lowest bit to one. This convention comes from LibVEX, and is not entirely our choice! If you see an odd ARM address, that just means the code at `address - 1` is in THUMB mode.

How do I serialize angr objects?

Pickle will work. However, python will default to using an extremely old pickle protocol that does not support more complex python data structures, so you must specify a more advanced data stream format. The easiest way to do this is `pickle.dumps(obj, -1)`.

What does `UnsupportedIROpError("floating point support disabled")` mean?

This might crop up if you're using a CGC analysis such as driller or rex. Floating point support in angr has been disabled in the CGC analyses for a tight-knit nebula of reasons:

- Libvex's representation of floating point numbers is imprecise - it converts the 80-bit extended precision format used by the x87 for computation to 64-bit doubles, making it impossible to get precise results
- There is very limited implementation support in angr for the actual primitive operations themselves as reported by libvex, so you will often get a less friendly "unsupported operation" error if you go too much further
- For what operations are implemented, the basic optimizations that allow tractability during symbolic computation (AST deduplication, operation collapsing) are not implemented for floating point ops, leading to gigantic ASTs
- There are memory corruption bugs in z3 that get triggered frighteningly easily when you're using huge workloads of mixed floating point and bitvector ops. We haven't been able to get a testcase that doesn't involve "just run angr" for the z3 guys to investigate.

Instead of trying to cope with all of these, we have simply disabled floating point support in the symbolic execution engine. To allow for execution in the presence of floating point ops, we have enabled an exploration technique called the https://github.com/angr/angr/blob/master/angr/exploration_techniques/oppologist.py that is supposed to catch these issues, concretize their inputs, and run the problematic instructions through qemu via unicorn engine, allowing execution to continue. The intuition is that the specific values of floating point operations don't typically affect the exploitation process.

If you're seeing this error and it's terminating the analysis, it's probably because you don't have unicorn installed or configured correctly. If you're seeing this issue just in a log somewhere, it's just the oppologist kicking in and you have nothing to worry about.

Why is angr's CFG different from IDA's?

Two main reasons:

- IDA does not split basic blocks at function calls. angr will, because they are a form of control flow and basic blocks end at control flow instructions. You generally do not need the supergraph for performing automated analyses.
- IDA will split basic blocks if another block jumps into the middle of it. This is called basic block normalization, and angr does not do it by default since it is unnecessary for most static analyses. You may enable it by passing `normalize=True` to the CFG analysis.

Why do I get incorrect register values when reading from a state during a SimInspect breakpoint?

libVEX will eliminate duplicate register writes within a single basic block when optimizations are enabled. Turn off IR optimization to make everything look right at all times.

In the case of the instruction pointer, libVEX will frequently omit mid-block writes even when optimizations are disabled. In this case, you should use `state.scratch.ins_addr` to get the current instruction pointer.

Core Concepts

Top Level Interfaces

Before You Start

Using and exploring angr in IPython (or other Python command line interpreters) is a main use case that we design angr for. When you are not sure what interfaces are available, tab completion is your friend!

Sometimes tab completion in IPython can be slow. We find the following workaround helpful without degrading the validity of completion results:

```
# Drop this file in IPython profile's startup directory to avoid running it every time.
import IPython
py = IPython.get_ipython()
py.Completer.use_jedi = False
```

Core Concepts

Before getting started with angr, you'll need to have a basic overview of some fundamental angr concepts and how to construct some basic angr objects. We'll go over this by examining what's directly available to you after you've loaded a binary!

Your first action with angr will always be to load a binary into a *project*. We'll use `/bin/true` for these examples.

```
>>> import angr
>>> proj = angr.Project('/bin/true')
```

A project is your control base in angr. With it, you will be able to dispatch analyses and simulations on the executable you just loaded. Almost every single object you work with in angr will depend on the existence of a project in some form.

Basic properties

First, we have some basic properties about the project: its CPU architecture, its filename, and the address of its entry point.

```
>>> import monkeyhex # this will format numerical results in hexadecimal
>>> proj.arch
<Arch AMD64 (LE)>
>>> proj.entry
0x401670
>>> proj.filename
'/bin/true'
```

- *arch* is an instance of an `archinfo.Arch` object for whichever architecture the program is compiled, in this case little-endian amd64. It contains a ton of clerical data about the CPU it runs on, which you can peruse at your leisure. The common ones you care about are `arch.bits`, `arch.bytes` (that one is a `@property` declaration on the main `Arch` class), `arch.name`, and `arch.memory_endness`.
- *entry* is the entry point of the binary!
- *filename* is the absolute filename of the binary. Riveting stuff!

The loader

Getting from a binary file to its representation in a virtual address space is pretty complicated! We have a module called CLE to handle that. CLE's result, called the loader, is available in the `.loader` property. We'll get into detail on how to use this soon, but for now just know that you can use it to see the shared libraries that angr loaded alongside your program and perform basic queries about the loaded address space.

```
>>> proj.loader
<Loaded true, maps [0x400000:0x5004000]>

>>> proj.loader.shared_objects # may look a little different for you!
{'ld-linux-x86-64.so.2': <ELF Object ld-2.24.so, maps [0x2000000:0x2227167]>,
 'libc.so.6': <ELF Object libc-2.24.so, maps [0x1000000:0x13c699f]>}

>>> proj.loader.min_addr
0x400000
>>> proj.loader.max_addr
0x5004000

>>> proj.loader.main_object  # we've loaded several binaries into this project. Here's the
<ELF Object true, maps [0x400000:0x60721f]>

```

```
>>> proj.loader.main_object.execstack  # sample query: does this binary have an executable
False
>>> proj.loader.main_object.pic  # sample query: is this binary position-independent?
True
```

The factory

There are a lot of classes in angr, and most of them require a project to be instantiated. Instead of making you pass around the project everywhere, we provide `project.factory`, which has several convenient constructors for common objects you'll want to use frequently.

This section will also serve as an introduction to several basic angr concepts. Strap in!

Blocks

First, we have `project.factory.block()`, which is used to extract a basic block of code from a given address. This is an important fact - *angr analyzes code in units of basic blocks.* You will get back a Block object, which can tell you lots of fun things about the block of code:

```
>>> block = proj.factory.block(proj.entry) # lift a block of code from the program's entry
<Block for 0x401670, 42 bytes>

>>> block.pp()                          # pretty-print a disassembly to stdout
0x401670:       xor     ebp, ebp
0x401672:       mov     r9, rdx
0x401675:       pop     rsi
0x401676:       mov     rdx, rsp
0x401679:       and     rsp, 0xfffffffffffffff0
0x40167d:       push    rax
0x40167e:       push    rsp
0x40167f:       lea     r8, [rip + 0x2e2a]
0x401686:       lea     rcx, [rip + 0x2db3]
0x40168d:       lea     rdi, [rip - 0xd4]
0x401694:       call    qword ptr [rip + 0x205866]

>>> block.instructions                  # how many instructions are there?
0xb
>>> block.instruction_addrs             # what are the addresses of the instructions?
[0x401670, 0x401672, 0x401675, 0x401676, 0x401679, 0x40167d, 0x40167e, 0x40167f, 0x401686,
```

Additionally, you can use a Block object to get other representations of the block of code:

```
>>> block.capstone                       # capstone disassembly
<CapstoneBlock for 0x401670>
>>> block.vex                            # VEX IRSB (that's a python internal address, not
<pyvex.block.IRSB at 0x7706330>
```

States

Here's another fact about angr - the `Project` object only represents an "initialization image" for the program. When you're performing execution with angr, you are working with a specific object representing a *simulated program state* - a `SimState`. Let's grab one right now!

```
>>> state = proj.factory.entry_state()
<SimState @ 0x401670>
```

A SimState contains a program's memory, registers, filesystem data... any "live data" that can be changed by execution has a home in the state. We'll cover how to interact with states in depth later, but for now, let's use `state.regs` and `state.mem` to access the registers and memory of this state:

```
>>> state.regs.rip        # get the current instruction pointer
<BV64 0x401670>
>>> state.regs.rax
<BV64 0x1c>
>>> state.mem[proj.entry].int.resolved  # interpret the memory at the entry point as a C i
<BV32 0x8949ed31>
```

Those aren't python ints! Those are *bitvectors*. Python integers don't have the same semantics as words on a CPU, e.g. wrapping on overflow, so we work with bitvectors, which you can think of as an integer as represented by a series of bits, to represent CPU data in angr. Note that each bitvector has a `.length` property describing how wide it is in bits.

We'll learn all about how to work with them soon, but for now, here's how to convert from python ints to bitvectors and back again:

```
>>> bv = state.solver.BVV(0x1234, 32)       # create a 32-bit-wide bitvector with value 0x
<BV32 0x1234>                               # BVV stands for bitvector value
>>> state.solver.eval(bv)                # convert to python int
0x1234
```

You can store these bitvectors back to registers and memory, or you can directly store a python integer and it'll be converted to a bitvector of the appropriate size:

```
>>> state.regs.rsi = state.solver.BVV(3, 64)
>>> state.regs.rsi
<BV64 0x3>

>>> state.mem[0x1000].long = 4
>>> state.mem[0x1000].long.resolved
<BV64 0x4>
```

The `mem` interface is a little confusing at first, since it's using some pretty hefty python magic. The short version of how to use it is:

- Use array[index] notation to specify an address
- Use `.<type>` to specify that the memory should be interpreted as <type> (common values: char, short,

int, long, size_t, uint8_t, uint16_t...)

- From there, you can either:
  - Store a value to it, either a bitvector or a python int
  - Use `.resolved` to get the value as a bitvector
  - Use `.concrete` to get the value as a python int

There are more advanced usages that will be covered later!

Finally, if you try reading some more registers you may encounter a very strange looking value:

```
>>> state.regs.rdi
<BV64 reg_48_11_64{UNINITIALIZED}>
```

This is still a 64-bit bitvector, but it doesn't contain a numerical value. Instead, it has a name! This is called a *symbolic variable* and it is the underpinning of symbolic execution. Don't panic! We will discuss all of this in detail exactly two chapters from now.

Simulation Managers

If a state lets us represent a program at a given point in time, there must be a way to get it to the *next* point in time. A simulation manager is the primary interface in angr for performing execution, simulation, whatever you want to call it, with states. As a brief introduction, let's show how to tick that state we created earlier forward a few basic blocks.

First, we create the simulation manager we're going to be using. The constructor can take a state or a list of states.

```
>>> simgr = proj.factory.simulation_manager(state)
<SimulationManager with 1 active>
>>> simgr.active
[<SimState @ 0x401670>]
```

A simulation manager can contain several *stashes* of states. The default stash, `active`, is initialized with the state we passed in. We could look at `simgr.active[0]` to look at our state some more, if we haven't had enough!

Now... get ready, we're going to do some execution.

```
>>> simgr.step()
```

We've just performed a basic block's worth of symbolic execution! We can look at the active stash again, noticing that it's been updated, and furthermore, that it has **not** modified our original state. SimState objects are treated as immutable by execution - you can safely use a single state as a "base" for multiple rounds of execution.

```
>>> simgr.active
```

```
[<SimState @ 0x1020300>]
>>> simgr.active[0].regs.rip                 # new and exciting!
<BV64 0x1020300>
>>> state.regs.rip                           # still the same!
<BV64 0x401670>
```

`/bin/true` isn't a very good example for describing how to do interesting things with symbolic execution, so we'll stop here for now.

Analyses

angr comes pre-packaged with several built-in analyses that you can use to extract some fun kinds of information from a program. Here they are:

```
>>> proj.analyses.            # Press TAB here in ipython to get an autocomplete-listing o
 proj.analyses.BackwardSlice        proj.analyses.CongruencyCheck      proj.analyses.reloa
 proj.analyses.BinaryOptimizer      proj.analyses.DDG                  proj.analyses.Stati
 proj.analyses.BinDiff              proj.analyses.DFG                  proj.analyses.Varial
 proj.analyses.BoyScout             proj.analyses.Disassembly          proj.analyses.Varial
 proj.analyses.CDG                  proj.analyses.GirlScout            proj.analyses.Verit
 proj.analyses.CFG                  proj.analyses.Identifier           proj.analyses.VFG
 proj.analyses.CFGEmulated          proj.analyses.LoopFinder           proj.analyses.VSA_DI
 proj.analyses.CFGFast              proj.analyses.Reassembler
```

A couple of these are documented later in this book, but in general, if you want to find how to use a given analysis, you should look in the api documentation. As an extremely brief example: here's how you construct and use a quick control-flow graph:

```
# Originally, when we loaded this binary it also loaded all its dependencies into the same
# This is undesirable for most analysis.
>>> proj = angr.Project('/bin/true', auto_load_libs=False)
>>> cfg = proj.analyses.CFGFast()
<CFGFast Analysis Result at 0x2d85130>

# cfg.graph is a networkx DiGraph full of CFGNode instances
# You should go look up the networkx APIs to learn how to use this!
>>> cfg.graph
<networkx.classes.digraph.DiGraph at 0x2da43a0>
>>> len(cfg.graph.nodes())
951

# To get the CFGNode for a given address, use cfg.get_any_node
>>> entry_node = cfg.get_any_node(proj.entry)
>>> len(list(cfg.graph.successors(entry_node)))
2
```

Now what?

Having read this page, you should now be aquainted with several important angr concepts: basic blocks, states, bitvectors, simulation managers, and analyses. You can't really do anything interesting besides just use angr as a glorified debugger, though! Keep reading, and you will unlock deeper powers...

Loading a Binary

Previously, you saw just the barest taste of angr's loading facilities - you loaded `/bin/true`, and then loaded it again without its shared libraries. You also saw `proj.loader` and a few things it could do. Now, we'll dive into the nuances of these interfaces and the things they can tell you.

We briefly mentioned angr's binary loading component, CLE. CLE stands for "CLE Loads Everything", and is responsible for taking a binary (and any libraries that it depends on) and presenting it to the rest of angr in a way that is easy to work with.

The Loader

Let's load `examples/fauxware/fauxware` and take a deeper look at how to interact with the loader.

```
>>> import angr, monkeyhex
>>> proj = angr.Project('examples/fauxware/fauxware')
>>> proj.loader
<Loaded fauxware, maps [0x400000:0x5008000]>
```

Loaded Objects

The CLE loader (`cle.Loader`) represents an entire conglomerate of loaded *binary objects*, loaded and mapped into a single memory space. Each binary object is loaded by a loader backend that can handle its filetype (a subclass of `cle.Backend`). For example, `cle.ELF` is used to load ELF binaries.

There will also be objects in memory that don't correspond to any loaded binary. For example, an object used to provide thread-local storage support, and an externs object used to provide unresolved symbols.

You can get the full list of objects that CLE has loaded with `loader.all_objects`, as well as several more targeted classifications:

```
# All loaded objects
>>> proj.loader.all_objects
[<ELF Object fauxware, maps [0x400000:0x60105f]>,
 <ELF Object libc-2.23.so, maps [0x1000000:0x13c999f]>,
 <ELF Object ld-2.23.so, maps [0x2000000:0x2227167]>,
 <ELFTLSObject Object cle##tls, maps [0x3000000:0x3015010]>,
 <ExternObject Object cle##externs, maps [0x4000000:0x4008000]>,
 <KernelObject Object cle##kernel, maps [0x5000000:0x5008000]>]

# This is the "main" object, the one that you directly specified when loading the project
>>> proj.loader.main_object
```

```
<ELF Object fauxware, maps [0x400000:0x60105f]>

# This is a dictionary mapping from shared object name to object
>>> proj.loader.shared_objects
{ 'fauxware': <ELF Object fauxware, maps [0x400000:0x60105f]>,
  'libc.so.6': <ELF Object libc-2.23.so, maps [0x1000000:0x13c999f]>,
  'ld-linux-x86-64.so.2': <ELF Object ld-2.23.so, maps [0x2000000:0x2227167]> }

# Here's all the objects that were loaded from ELF files
# If this were a windows program we'd use all_pe_objects!
>>> proj.loader.all_elf_objects
[<ELF Object fauxware, maps [0x400000:0x60105f]>,
 <ELF Object libc-2.23.so, maps [0x1000000:0x13c999f]>,
 <ELF Object ld-2.23.so, maps [0x2000000:0x2227167]>]

# Here's the "externs object", which we use to provide addresses for unresolved imports an
>>> proj.loader.extern_object
<ExternObject Object cle##externs, maps [0x4000000:0x4008000]>

# This object is used to provide addresses for emulated syscalls
>>> proj.loader.kernel_object
<KernelObject Object cle##kernel, maps [0x5000000:0x5008000]>

# Finally, you can to get a reference to an object given an address in it
>>> proj.loader.find_object_containing(0x400000)
<ELF Object fauxware, maps [0x400000:0x60105f]>
```

You can interact directly with these objects to extract metadata from them:

```
>>> obj = proj.loader.main_object

# The entry point of the object
>>> obj.entry
0x400580

>>> obj.min_addr, obj.max_addr
(0x400000, 0x60105f)

# Retrieve this ELF's segments and sections
>>> obj.segments
<Regions: [<ELFSegment memsize=0xa74, filesize=0xa74, vaddr=0x400000, flags=0x5, offset=0x
           <ELFSegment memsize=0x238, filesize=0x228, vaddr=0x600e28, flags=0x6, offset=0x
>>> obj.sections
<Regions: [<Unnamed | offset 0x0, vaddr 0x0, size 0x0>,
           <.interp | offset 0x238, vaddr 0x400238, size 0x1c>,
           <.note.ABI-tag | offset 0x254, vaddr 0x400254, size 0x20>,
            ...etc

# You can get an individual segment or section by an address it contains:
>>> obj.find_segment_containing(obj.entry)
<ELFSegment memsize=0xa74, filesize=0xa74, vaddr=0x400000, flags=0x5, offset=0x0>
>>> obj.find_section_containing(obj.entry)
```

```
<.text | offset 0x580, vaddr 0x400580, size 0x338>

# Get the address of the PLT stub for a symbol
>>> addr = obj.plt['strcmp']
>>> addr
0x400550
>>> obj.reverse_plt[addr]
'strcmp'

# Show the prelinked base of the object and the location it was actually mapped into memory
>>> obj.linked_base
0x400000
>>> obj.mapped_base
0x400000
```

Symbols and Relocations

You can also work with symbols while using CLE. A symbol is a fundamental concept in the world of executable formats, effectively mapping a name to an address.

The easiest way to get a symbol from CLE is to use `loader.find_symbol`, which takes either a name or an address and returns a Symbol object.

```
>>> strcmp = proj.loader.find_symbol('strcmp')
>>> strcmp
<Symbol "strcmp" in libc.so.6 at 0x1089cd0>
```

The most useful attributes on a symbol are its name, its owner, and its address, but the "address" of a symbol can be ambiguous. The Symbol object has three ways of reporting its address:

- `.rebased_addr` is its address in the global address space. This is what is shown in the print output.
- `.linked_addr` is its address relative to the prelinked base of the binary. This is the address reported in, for example, `readelf(1)`.
- `.relative_addr` is its address relative to the object base. This is known in the literature (particularly the Windows literature) as an RVA (relative virtual address).

```
>>> strcmp.name
'strcmp'

>>> strcmp.owner
<ELF Object libc-2.23.so, maps [0x1000000:0x13c999f]>

>>> strcmp.rebased_addr
0x1089cd0
>>> strcmp.linked_addr
0x89cd0
>>> strcmp.relative_addr
0x89cd0
```

In addition to providing debug information, symbols also support the notion of dynamic linking. libc provides the strcmp symbol as an export, and the main binary depends on it. If we ask CLE to give us a strcmp symbol from the main object directly, it'll tell us that this is an *import symbol*. Import symbols do not have meaningful addresses associated with them, but they do provide a reference to the symbol that was used to resolve them, as `.resolvedby`.

```
>>> strcmp.is_export
True
>>> strcmp.is_import
False

# On Loader, the method is find_symbol because it performs a search operation to find the 
# On an individual object, the method is get_symbol because there can only be one symbol w
>>> main_strcmp = proj.loader.main_object.get_symbol('strcmp')
>>> main_strcmp
<Symbol "strcmp" in fauxware (import)>
>>> main_strcmp.is_export
False
>>> main_strcmp.is_import
True
>>> main_strcmp.resolvedby
<Symbol "strcmp" in libc.so.6 at 0x1089cd0>
```

The specific ways that the links between imports and exports should be registered in memory are handled by another notion called *relocations*. A relocation says, "when you match *[import]* up with an export symbol, please write the export's address to *[location]*, formatted as *[format]*." We can see the full list of relocations for an object (as `Relocation` instances) as `obj.relocs`, or just a mapping from symbol name to Relocation as `obj.imports`. There is no corresponding list of export symbols.

A relocation's corresponding import symbol can be accessed as `.symbol`. The address the relocation will write to is accessable through any of the address identifiers you can use for Symbol, and you can get a reference to the object requesting the relocation with `.owner` as well.

```
# Relocations don't have a good pretty-printing, so those addresses are python-internal, u
>>> proj.loader.shared_objects['libc.so.6'].imports
{'__libc_enable_secure': <cle.backends.elf.relocation.amd64.R_X86_64_GLOB_DAT at 0x7ff5c5f
 '__tls_get_addr': <cle.backends.elf.relocation.amd64.R_X86_64_JUMP_SLOT at 0x7ff5c601835
 '_dl_argv': <cle.backends.elf.relocation.amd64.R_X86_64_GLOB_DAT at 0x7ff5c5fd2e48>,
 '_dl_find_dso_for_object': <cle.backends.elf.relocation.amd64.R_X86_64_JUMP_SLOT at 0x7ff
 '_dl_starting_up': <cle.backends.elf.relocation.amd64.R_X86_64_GLOB_DAT at 0x7ff5c5fd2550
 '_rtld_global': <cle.backends.elf.relocation.amd64.R_X86_64_GLOB_DAT at 0x7ff5c5fce4e0>,
 '_rtld_global_ro': <cle.backends.elf.relocation.amd64.R_X86_64_GLOB_DAT at 0x7ff5c5fcea20
```

If an import cannot be resolved to any export, for example, because a shared library could not be found, CLE will automatically update the externs object (`loader.extern_obj`) to claim it provides the symbol as an export.

Loading Options

If you are loading something with `angr.Project` and you want to pass an option to the `cle.Loader` instance that Project implicitly creates, you can just pass the keyword argument directly to the Project constructor, and it will be passed on to CLE. You should look at the CLE API docs. if you want to know everything that could possibly be passed in as an option, but we will go over some important and frequently used options here.

Basic Options

We've discussed `auto_load_libs` already - it enables or disables CLE's attempt to automatically resolve shared library dependencies, and is on by default. Additionally, there is the opposite, `except_missing_libs`, which, if set to true, will cause an exception to be thrown whenever a binary has a shared library dependency that cannot be resolved.

You can pass a list of strings to `force_load_libs` and anything listed will be treated as an unresolved shared library dependency right out of the gate, or you can pass a list of strings to `skip_libs` to prevent any library of that name from being resolved as a dependency. Additionally, you can pass a list of strings (or a single string) to `ld_path`, which will be used as an additional search path for shared libraries, before any of the defaults: the same directory as the loaded program, the current working directory, and your system libraries.

Per-Binary Options

If you want to specify some options that only apply to a specific binary object, CLE will let you do that too. The parameters `main_opts` and `lib_opts` do this by taking dictionaries of options. `main_opts` is a mapping from option names to option values, while `lib_opts` is a mapping from library name to dictionaries mapping option names to option values.

The options that you can use vary from backend to backend, but some common ones are:

- `backend` - which backend to use, as either a class or a name
- `base_addr` - a base address to use
- `entry_point` - an entry point to use
- `arch` - the name of an architecture to use

Example:

```
>>> angr.Project('examples/fauxware/fauxware', main_opts={'backend': 'blob', 'arch': 'i386
<Project examples/fauxware/fauxware>
```

Backends

CLE currently has backends for statically loading ELF, PE, CGC, Mach-O and ELF core dump files, as well as loading files into a flat address space. CLE will automatically detect the correct backend to use in most cases, so you shouldn't need to specify which backend you're using unless you're doing some pretty weird stuff.

You can force CLE to use a specific backend for an object by including a key in its options dictionary, as described above. Some backends cannot autodetect which architecture to use and *must* have a `arch` specified. The key doesn't need to match any list of architectures; angr will identify which architecture you mean given almost any common identifier for any supported arch.

To refer to a backend, use the name from this table:

backend name	description	requires `arch`?
elf	Static loader for ELF files based on PyELFTools	no
pe	Static loader for PE files based on PEFile	no
mach-o	Static loader for Mach-O files. Does not support dynamic linking or rebasing.	no
cgc	Static loader for Cyber Grand Challenge binaries	no
backedcgc	Static loader for CGC binaries that allows specifying memory and register backers	no
elfcore	Static loader for ELF core dumps	no
blob	Loads the file into memory as a flat image	yes

Symbolic Function Summaries

By default, Project tries to replace external calls to library functions by using symbolic summaries termed *SimProcedures* - effectively just python functions that imitate the library function's effect on the state. We've implemented a whole bunch of functions as SimProcedures. These builtin procedures are available in the `angr.SIM_PROCEDURES` dictionary, which is two-leveled, keyed first on the package name (libc, posix, win32, stubs) and then on the name of the library function. Executing a SimProcedure instead of the actual library function that gets loaded from your system makes analysis a LOT more tractable, at the cost of some potential inaccuracies.

When no such summary is available for a given function:

- if `auto_load_libs` is `True` (this is the default), then the *real* library function is executed instead. This may or may not be what you want, depending on the actual function. For example, some of libc's functions are extremely complex to analyze and will most likely cause an explosion of the number of states for the path trying to execute them.
- if `auto_load_libs` is `False`, then external functions are unresolved, and Project will resolve them to a generic "stub" SimProcedure called `ReturnUnconstrained`. It does what its name says: it returns a unique unconstrained symbolic value each time it is called.
- if `use_sim_procedures` (this is a parameter to `angr.Project`, not `cle.Loader`) is `False` (it is `True` by default), then only symbols provided by the extern object will be replaced with SimProcedures, and they will be replaced by a stub `ReturnUnconstrained`, which does nothing but return a symbolic value.
- you may specify specific symbols to exclude from being replaced with SimProcedures with the parameters to `angr.Project`: `exclude_sim_procedures_list` and `exclude_sim_procedures_func`.
- Look at the code for `angr.Project._register_object` for the exact algorithm.

Hooking

The mechanism by which angr replaces library code with a python summary is called hooking, and you can do it too! When performing simulation, at every step angr checks if the current address has been hooked, and if so, runs the hook instead of the binary code at that address. The API to let you do this is `proj.hook(addr, hook)`, where `hook` is a SimProcedure instance. You can manage your project's hooks with `.is_hooked`, `.unhook`, and `.hooked_by`, which should hopefully not require explanation.

There is an alternate API for hooking an address that lets you specify your own off-the-cuff function to use as a hook, by using `proj.hook(addr)` as a function decorator. If you do this, you can also optionally specify a `length` keyword argument to make execution jump some number of bytes forward after your hook finishes.

```
>>> stub_func = angr.SIM_PROCEDURES['stubs']['ReturnUnconstrained'] # this is a CLASS
>>> proj.hook(0x10000, stub_func())  # hook with an instance of the class

>>> proj.is_hooked(0x10000)            # these functions should be pretty self-explanitory
True
>>> proj.hooked_by(0x10000)
<ReturnUnconstrained>
>>> proj.unhook(0x10000)

>>> @proj.hook(0x20000, length=5)
... def my_hook(state):
...     state.regs.rax = 1
```

```

>>> proj.is_hooked(0x20000)
True
```

Furthermore, you can use `proj.hook_symbol(name, hook)`, providing the name of a symbol as the first argument, to hook the address where the symbol lives. One very important usage of this is to extend the behavior of angr's built-in library SimProcedures. Since these library functions are just classes, you can subclass them, overriding pieces of their behavior, and then use your subclass in a hook.

So far so good!

By now, you should have a reasonable understanding of how to control the environment in which your analysis happens, on the level of the CLE loader and the angr Project. You should also understand that angr makes a reasonable attempt to simplify its analysis by hooking complex library functions with SimProcedures that summarize the effects of the functions.

In order to see all the things you can do with the CLE loader and its backends, look at the CLE API docs.

Solver Engine

angr's power comes not from it being an emulator, but from being able to execute with what we call *symbolic variables*. Instead of saying that a variable has a *concrete* numerical value, we can say that it holds a *symbol*, effectively just a name. Then, performing arithmetic operations with that variable will yield a tree of operations (termed an *abstract syntax tree* or *AST*, from compiler theory). ASTs can be translated into constraints for an *SMT solver*, like z3, in order to ask questions like *"given the output of this sequence of operations, what must the input have been?"* Here, you'll learn how to use angr to answer this.

Working with Bitvectors

Let's get a dummy project and state so we can start playing with numbers.

```
>>> import angr, monkeyhex
>>> proj = angr.Project('/bin/true')
>>> state = proj.factory.entry_state()
```

A bitvector is just a sequence of bits, interpreted with the semantics of a bounded integer for arithmetic. Let's make a few.

```
# 64-bit bitvectors with concrete values 1 and 100
>>> one = state.solver.BVV(1, 64)
>>> one
```

```
 <BV64 0x1>
>>> one_hundred = state.solver.BVV(100, 64)
>>> one_hundred
 <BV64 0x64>

# create a 27-bit bitvector with concrete value 9
>>> weird_nine = state.solver.BVV(9, 27)
>>> weird_nine
<BV27 0x9>
```

As you can see, you can have any sequence of bits and call them a bitvector. You can do math with them too:

```
>>> one + one_hundred
<BV64 0x65>

# You can provide normal python integers and they will be coerced to the appropriate type:
>>> one_hundred + 0x100
<BV64 0x164>

# The semantics of normal wrapping arithmetic apply
>>> one_hundred - one*200
<BV64 0xffffffffffffff9c>
```

You *cannot* say `one + weird_nine`, though. It is a type error to perform an operation on bitvectors of differing lengths. You can, however, extend `weird_nine` so it has an appropriate number of bits:

```
>>> weird_nine.zero_extend(64 - 27)
<BV64 0x9>
>>> one + weird_nine.zero_extend(64 - 27)
<BV64 0xa>
```

`zero_extend` will pad the bitvector on the left with the given number of zero bits. You can also use `sign_extend` to pad with a duplicate of the highest bit, preserving the value of the bitvector under two's compliment signed integer semantics.

Now, let's introduce some symbols into the mix.

```
# Create a bitvector symbol named "x" of length 64 bits
>>> x = state.solver.BVS("x", 64)
>>> x
<BV64 x_9_64>
>>> y = state.solver.BVS("y", 64)
>>> y
<BV64 y_10_64>
```

`x` and `y` are now *symbolic variables*, which are kind of like the variables you learned to work with in 7th grade algebra. Notice that the name you provided has been been mangled by appending an incrementing counter and You can do as much arithmetic as you want with them, but you won't get a number back, you'll get an AST instead.

```
>>> x + one
<BV64 x_9_64 + 0x1>

>>> (x + one) / 2
<BV64 (x_9_64 + 0x1) / 0x2>

>>> x - y
<BV64 x_9_64 - y_10_64>
```

Technically `x` and `y` and even `one` are also ASTs - any bitvector is a tree of operations, even if that tree is only one layer deep. To understand this, let's learn how to process ASTs.

Each AST has a `.op` and a `.args`. The op is a string naming the operation being performed, and the args are the values the operation takes as input. Unless the op is `BVV` or `BVS` (or a few others...), the args are all other ASTs, the tree eventually terminating with BVVs or BVSs.

```
>>> tree = (x + 1) / (y + 2)
>>> tree
<BV64 (x_9_64 + 0x1) / (y_10_64 + 0x2)>
>>> tree.op
'__floordiv__'
>>> tree.args
(<BV64 x_9_64 + 0x1>, <BV64 y_10_64 + 0x2>)
>>> tree.args[0].op
'__add__'
>>> tree.args[0].args
(<BV64 x_9_64>, <BV64 0x1>)
>>> tree.args[0].args[1].op
'BVV'
>>> tree.args[0].args[1].args
(1, 64)
```

From here on out, we will use the word "bitvector" to refer to any AST whose topmost operation produces a bitvector. There can be other data types represented through ASTs, including floating point numbers and, as we're about to see, booleans.

Symbolic Constraints

Performing comparison operations between any two similarly-typed ASTs will yield another AST - not a bitvector, but now a symbolic boolean.

```
>>> x == 1
<Bool x_9_64 == 0x1>
>>> x == one
```

```
<Bool x_9_64 == 0x1>
>>> x > 2
<Bool x_9_64 > 0x2>
>>> x + y == one_hundred + 5
<Bool (x_9_64 + y_10_64) == 0x69>
>>> one_hundred > 5
<Bool True>
>>> one_hundred > -5
<Bool False>
```

One tidbit you can see from this is that the comparisons are unsigned by default. The -5 in the last example is coerced to `<BV64 0xfffffffffffffffb>`, which is definitely not less than one hundred. If you want the comparison to be signed, you can say `one_hundred.SGT(-5)` (that's "signed greater-than"). A full list of operations can be found at the end of this chapter.

This snippet also illustrates an important point about working with angr - you should never directly use a comparison between variables in the condition for an if- or while-statement, since the answer might not have a concrete truth value. Even if there is a concrete truth value, `if one > one_hundred` will raise an exception. Instead, you should use `solver.is_true` and `solver.is_false`, which test for concrete truthyness/falsiness without performing a constraint solve.

```
>>> yes = one == 1
>>> no = one == 2
>>> maybe = x == y
>>> state.solver.is_true(yes)
True
>>> state.solver.is_false(yes)
False
>>> state.solver.is_true(no)
False
>>> state.solver.is_false(no)
True
>>> state.solver.is_true(maybe)
False
>>> state.solver.is_false(maybe)
False
```

Constraint Solving

You can treat any symbolic boolean as an assertion about the valid values of a symbolic variable by adding it as a *constraint* to the state. You can then query for a valid value of a symbolic variable by asking for an evaluation of a symbolic expression.

An example will probably be more clear than an explanation here:

```
>>> state.solver.add(x > y)
>>> state.solver.add(y > 2)
>>> state.solver.add(10 > x)
>>> state.solver.eval(x)
4
```

By adding these constraints to the state, we've forced the constraint solver to consider them as assertions that must be satisfied about any values it returns. If you run this code, you might get a different value for x, but that value will definitely be greater than 3 (since y must be greater than 2 and x must be greater than y) and less than 10. Furthermore, if you then say `state.solver.eval(y)`, you'll get a value of y which is consistent with the value of x that you got. If you don't add any constraints between two queries, the results will be consistent with each other.

From here, it's easy to see how to do the task we proposed at the beginning of the chapter - finding the input that produced a given output.

```
# get a fresh state without constraints
>>> state = proj.factory.entry_state()
>>> input = state.solver.BVS('input', 64)
>>> operation = (((input + 4) * 3) >> 1) + input
>>> output = 200
>>> state.solver.add(operation == output)
>>> state.solver.eval(input)
0x3333333333333381
```

Note that, again, this solution only works because of the bitvector semantics. If we were operating over the domain of integers, there would be no solutions!

If we add conflicting or contradictory constraints, such that there are no values that can be assigned to the variables such that the constraints are satisfied, the state becomes *unsatisfiable*, or unsat, and queries against it will raise an exception. You can check the satisfiability of a state with `state.satisfiable()`.

```
>>> state.solver.add(input < 2**32)
>>> state.satisfiable()
False
```

You can also evaluate more complex expressions, not just single variables.

```
# fresh state
>>> state = proj.factory.entry_state()
>>> state.solver.add(x - y >= 4)
>>> state.solver.add(y > 0)
>>> state.solver.eval(x)
5
>>> state.solver.eval(y)
1
>>> state.solver.eval(x + y)
6
```

From this we can see that `eval` is a general purpose method to convert any bitvector into a python primitive while respecting the integrity of the state. This is why we use `eval` to convert from concrete bitvectors to python ints, too!

Also note that the x and y variables can be used in this new state despite having been created using an old state. Variables are not tied to any one state, and can exist freely.

Floating point numbers

z3 has support for the theory of IEEE754 floating point numbers, and so angr can use them as well. The main difference is that instead of a width, a floating point number has a *sort*. You can create floating point symbols and values with `FPV` and `FPS`.

```
# fresh state
>>> state = proj.factory.entry_state()
>>> a = state.solver.FPV(3.2, state.solver.fp.FSORT_DOUBLE)
>>> a
<FP64 FPV(3.2, DOUBLE)>

>>> b = state.solver.FPS('b', state.solver.fp.FSORT_DOUBLE)
>>> b
<FP64 FPS('FP_b_0_64', DOUBLE)>

>>> a + b
<FP64 fpAdd('RNE', FPV(3.2, DOUBLE), FPS('FP_b_0_64', DOUBLE))>

>>> a + 4.4
<FP64 FPV(7.6000000000000005, DOUBLE)>

>>> b + 2 < 0
<Bool fpLT(fpAdd('RNE', FPS('FP_b_0_64', DOUBLE), FPV(2.0, DOUBLE)), FPV(0.0, DOUBLE))>
```

So there's a bit to unpack here - for starters the pretty-printing isn't as smart about floating point numbers. But past that, most operations actually have a third parameter, implicitly added when you use the binary operators - the rounding mode. The IEEE754 spec supports multiple rounding modes (round-to-nearest, round-to-zero, round-to-positive, etc), so z3 has to support them. If you want to specify the rounding mode for an operation, use the fp operation explicitly (`solver.fpAdd` for example) with a rounding mode (one of `solver.fp.RM_*`) as the first argument.

Constraints and solving work in the same way, but with `eval` returning a floating point number:

```
>>> state.solver.add(b + 2 < 0)
>>> state.solver.add(b + 2 > -1)
>>> state.solver.eval(b)
-2.4999999999999996
```

This is nice, but sometimes we need to be able to work directly with the representation of the float as a bitvector. You can interpret bitvectors as floats and vice versa, with the methods `raw_to_bv` and `raw_to_fp`:

```
>>> a.raw_to_bv()
<BV64 0x400999999999999a>
>>> b.raw_to_bv()
<BV64 fpToIEEEBV(FPS('FP_b_0_64', DOUBLE))>

>>> state.solver.BVV(0, 64).raw_to_fp()
<FP64 FPV(0.0, DOUBLE)>
>>> state.solver.BVS('x', 64).raw_to_fp()
<FP64 fpToFP(x_1_64, DOUBLE)>
```

These conversions preserve the bit-pattern, as if you casted a float pointer to an int pointer or vice versa. However, if you want to preserve the value as closely as possible, as if you casted a float to an int (or vice versa), you can use a different set of methods, `val_to_fp` and `val_to_bv`. These methods must take the size or sort of the target value as a parameter, due to the floating-point nature of floats.

```
>>> a
<FP64 FPV(3.2, DOUBLE)>
>>> a.val_to_bv(12)
<BV12 0x3>
>>> a.val_to_bv(12).val_to_fp(state.solver.fp.FSORT_FLOAT)
<FP32 FPV(3.0, FLOAT)>
```

These methods can also take a `signed` parameter, designating the signedness of the source or target bitvector.

More Solving Methods

`eval` will give you one possible solution to an expression, but what if you want several? What if you want to ensure that the solution is unique? The solver provides you with several methods for common solving patterns:

- `solver.eval(expression)` will give you one possible solution to the given expression.
- `solver.eval_one(expression)` will give you the solution to the given expression, or throw an error if more than one solution is possible.
- `solver.eval_upto(expression, n)` will give you up to n solutions to the given expression, returning fewer than n if fewer than n are possible.
- `solver.eval_atleast(expression, n)` will give you n solutions to the given expression, throwing an error if fewer than n are possible.
- `solver.eval_exact(expression, n)` will give you n solutions to the given expression, throwing an error if fewer or more than are possible.
- `solver.min(expression)` will give you the minimum possible solution to the given expression.
- `solver.max(expression)` will give you the maximum possible solution to the given expression.

Additionally, all of these methods can take the following keyword arguments:

- `extra_constraints` can be passed as a tuple of constraints.
  These constraints will be taken into account for this evaluation, but will not be added to the state.
- `cast_to` can be passed a data type to cast the result to.
  Currently, this can only be `int` and `bytes`, which will cause the method to return the corresponding representation of the underlying data.
  For example, `state.solver.eval(state.solver.BVV(0x41424344, 32), cast_to=bytes)` will return `b'ABCD'`.

Summary

That was a lot!! After reading this, you should be able to create and manipulate bitvectors, booleans, and floating point values to form trees of operations, and then query the constraint solver attached to a state for possible solutions under a set of constraints. Hopefully by this point you understand the power of using ASTs to represent computations, and the power of a constraint solver.

In the appendix, you can find a reference for all the additional operations you can apply to ASTs, in case you ever need a quick table to look at.

Program State

So far, we've only used angr's simulated program states (`SimState` objects) in the barest possible way in order to demonstrate basic concepts about angr's operation. Here, you'll learn about the structure of a state object and how to interact with it in a variety of useful ways.

Review: Reading and writing memory and registers

If you've been reading this book in order (and you should be, at least for this first section), you already saw the basics of how to access memory and registers. `state.regs` provides read and write access to the registers through attributes with the names of each register, and `state.mem` provides typed read and write access to memory with index-access notation to specify the address followed by an attribute access to specify the type you would like to interpret the memory as.

Additionally, you should now know how to work with ASTs, so you can now understand that any bitvector-typed AST can be stored in registers or memory.

Here are some quick examples for copying and performing operations on data from the state:

```
>>> import angr, claripy
>>> proj = angr.Project('/bin/true')
>>> state = proj.factory.entry_state()
# copy rsp to rbp
>>> state.regs.rbp = state.regs.rsp

# store rdx to memory at 0x1000
>>> state.mem[0x1000].uint64_t = state.regs.rdx

# dereference rbp
>>> state.regs.rbp = state.mem[state.regs.rbp].uint64_t.resolved

# add rax, qword ptr [rsp + 8]
>>> state.regs.rax += state.mem[state.regs.rsp + 8].uint64_t.resolved
```

Basic Execution

Earlier, we showed how to use a Simulation Manager to do some basic execution. We'll show off the full capabilities of the simulation manager in the next chapter, but for now we can use a much simpler interface to demonstrate how symbolic execution works: `state.step()`. This method will perform one step of symbolic execution and return an object called `SimSuccessors`. Unlike normal emulation, symbolic execution can produce several successor states that can be classified in a number of ways. For now, what we care about is the `.successors` property of this object, which is a list containing all the "normal" successors of a given step.

Why a list, instead of just a single successor state? Well, angr's process of symbolic execution is just the taking the operations of the individual instructions compiled into the program and performing them to mutate a SimState. When a line of code like `if (x > 4)` is reached, what happens if x is a symbolic bitvector? Somewhere in the depths of angr, the comparison `x > 4` is going to get performed, and the result is going to be `<Bool x_32_1 > 4>`.

That's fine, but the next question is, do we take the "true" branch or the "false" one? The answer is, we take both! We generate two entirely separate successor states - one simulating the case where the condition was true and simulating the case where the condition was false. In the first state, we add `x > 4` as a constraint, and in the second state, we add `!(x > 4)` as a constraint. That way, whenever we perform a constraint solve using either of these successor states, *the conditions on the state ensure that any solutions we get are valid inputs that will cause execution to follow the same path that the given state has followed.*

To demonstrate this, let's use a fake firmware image as an example. If you look at the source code for this binary, you'll see that the authentication mechanism for the firmware is backdoored; any username can be authenticated as an administrator with the password "SOSNEAKY". Furthermore, the first comparison against user input that happens is the comparison against the backdoor, so if we step until we get more than one successor state, one of those states will contain conditions constraining the user input to be the backdoor password. The following snippet implements this:

```
>>> proj = angr.Project('examples/fauxware/fauxware')
>>> state = proj.factory.entry_state(stdin=angr.SimFile)  # ignore that argument for now -
>>> while True:
...     succ = state.step()
...     if len(succ.successors) == 2:
...         break
...     state = succ.successors[0]

>>> state1, state2 = succ.successors
>>> state1
<SimState @ 0x400629>
>>> state2
<SimState @ 0x400699
```

Don't look at the constraints on these states directly - the branch we just went through involves the result of `strcmp`, which is a tricky function to emulate symbolically, and the resulting constraints are *very* complicated.

The program we emulated took data from standard input, which angr treats as an infinite stream of symbolic data by default. To perform a constraint solve and get a possible value that input could have taken in order to satisfy the constraints, we'll need to get a reference to the actual contents of stdin. We'll go over how our file and input subsystems work later on this very page, but for now, just use `state.posix.stdin.load(0, state.posix.stdin.size)` to retrieve a bitvector representing all the content read from stdin so far.

```
>>> input_data = state1.posix.stdin.load(0, state.posix.stdin.size)

>>> state1.solver.eval(input_data, cast_to=bytes)
b'\x00\x00\x00\x00\x00\x00\x00\x00\x00SOSNEAKY\x00\x00\x00'

>>> state2.solver.eval(input_data, cast_to=bytes)

b'\x00\x00\x00\x00\x00\x00\x00\x00\x00S\x00\x80N\x00\x00 \x00\x00\x00\x00'
```

As you can see, in order to go down the `state1` path, you must have given as a password the backdoor string "SOSNEAKY". In order to go down the `state2` path, you must have given something *besides* "SOSNEAKY". z3 has helpfully provided one of the billions of strings fitting this criteria.

Fauxware was the first program angr's symbolic execution ever successfully worked on, back in 2013. By finding its backdoor using angr you are participating in a grand tradition of having a bare-bones understanding of how to use symbolic execution to extract meaning from binaries!

State Presets

So far, whenever we've been working with a state, we've created it with `project.factory.entry_state()`. This is just one of several *state constructors* available on the project factory:

- `.blank_state()` constructs a "blank slate" blank state, with most of its data left uninitialized. When accessing uninitialized data, an unconstrained symbolic value will be returned.
- `.entry_state()` constructs a state ready to execute at the main binary's entry point.
- `.full_init_state()` constructs a state that is ready to execute through any initializers that need to

be run before the main binary's entry point, for example, shared library constructors or preinitializers. When it is finished with these it will jump to the entry point.

- `.call_state()` constructs a state ready to execute a given function.

You can customize the state through several arguments to these constructors:

- All of these constructors can take an `addr` argument to specify the exact address to start.
- If you're executing in an environment that can take command line arguments or an environment, you can pass a list of arguments through `args` and a dictionary of environment variables through `env` into `entry_state` and `full_init_state`. The values in these structures can be strings or bitvectors, and will be serialized into the state as the arguments and environment to the simulated execution. The default `args` is an empty list, so if the program you're analyzing expects to find at least an `argv[0]`, you should always provide that!
- If you'd like to have `argc` be symbolic, you can pass a symbolic bitvector as `argc` to the `entry_state` and `full_init_state` constructors. Be careful, though: if you do this, you should also add a constraint to the resulting state that your value for argc cannot be larger than the number of args you passed into `args`.
- To use the call state, you should call it with `.call_state(addr, arg1, arg2, ...)`, where `addr` is the address of the function you want to call and `argN` is the Nth argument to that function, either as a python integer, string, or array, or a bitvector. If you want to have memory allocated and actually pass in a pointer to an object, you should wrap it in an PointerWrapper, i.e. `angr.PointerWrapper("point to me!")`. The results of this API can be a little unpredictable, but we're working on it.

- To specify the calling convention used for a function with `call_state`, you can pass a `SimCC` instance as the `cc` argument.
  We try to pick a sane default, but for special cases you will need to help angr out.

There are several more options that can be used in any of these constructors! See the docs on the `project.factory` object (an `AngrObjectFactory`) for more details.

Low level interface for memory

The `state.mem` interface is convenient for loading typed data from memory, but when you want to do raw loads and stores to and from ranges of memory, it's very cumbersome. It turns out that `state.mem` is actually just a bunch of logic to correctly access the underlying memory storage, which is just a flat address space filled with bitvector data: `state.memory`. You can use `state.memory` directly with the `.load(addr, size)` and `.store(addr, val)` methods:

```
>>> s = proj.factory.blank_state()
>>> s.memory.store(0x4000, s.solver.BVV(0x0123456789abcdef0123456789abcdef, 128))
>>> s.memory.load(0x4004, 6) # load-size is in bytes
<BV48 0x89abcdef0123>
```

As you can see, the data is loaded and stored in a "big-endian" fashion, since the primary purpose of `state.memory` is to load an store swaths of data with no attached semantics. However, if you want to perform a byteswap on the loaded or stored data, you can pass a keyword argument `endness` - if you specify little-endian, byteswap will happen. The endness should be one of the members of the `Endness` enum in the `archinfo` package used to hold declarative data about CPU architectures for angr. Additionally, the endness of the program being analyzed can be found as `arch.memory_endness` - for instance `state.arch.memory_endness`.

```
>>> import archinfo
>>> s.memory.load(0x4000, 4, endness=archinfo.Endness.LE)
<BV32 0x67452301>
```

There is also a low-level interface for register access, `state.registers`, that uses the exact same API as `state.memory`, but explaining its behavior involves a dive into the abstractions that angr uses to seamlessly work with multiple architectures. The short version is that it is simply a register file, with the mapping between registers and offsets defined in archinfo.

State Options

There are a lot of little tweaks that can be made to the internals of angr that will optimize behavior in some situations and be a detriment in others. These tweaks are controlled through state options.

On each SimState object, there is a set (`state.options`) of all its enabled options. Each option (really just a string) controls the behavior of angr's execution engine in some minute way. A listing of the full domain of options, along with the defaults for different state types, can be found in the appendix. You can access an individual option for adding to a state through `angr.options`. The individual options are named with CAPITAL_LETTERS, but there are also common groupings of objects that you might want to use bundled together, named with lowercase_letters.

When creating a SimState through any constructor, you may pass the keyword arguments `add_options` and `remove_options`, which should be sets of options that modify the initial options set from the default.

```
# Example: enable lazy solves, an option that causes state satisfiability to be checked as
# This change to the settings will be propagated to all successor states created from this
>>> s.options.add(angr.options.LAZY_SOLVES)

# Create a new state with lazy solves enabled
>>> s = proj.factory.entry_state(add_options={angr.options.LAZY_SOLVES})

# Create a new state without simplification options enabled
>>> s = proj.factory.entry_state(remove_options=angr.options.simplification)
```

State Plugins

With the exception of the set of options just discussed, everything stored in a SimState is actually stored in a *plugin* attached to the state. Almost every property on the state we've discussed so far is a plugin - `memory`, `registers`, `mem`, `regs`, `solver`, etc. This design allows for code modularity as well as the ability to easily implement new kinds of data storage for other aspects of an emulated state, or the ability to provide alternate implementations of plugins.

For example, the normal `memory` plugin simulates a flat memory space, but analyses can choose to enable the "abstract memory" plugin, which uses alternate data types for addresses to simulate free-floating memory mappings independent of address, to provide `state.memory`. Conversely, plugins can reduce code complexity: `state.memory` and `state.registers` are actually two different instances of the same plugin, since the registers are emulated with an address space as well.

The globals plugin

`state.globals` is an extremely simple plugin: it implements the interface of a standard python dict, allowing you to store arbitrary data on a state.

The history plugin

`state.history` is a very important plugin storing historical data about the path a state has taken during execution. It is actually a linked list of several history nodes, each one representing a single round of execution---you can traverse this list with `state.history.parent.parent` etc.

To make it more convenient to work with this structure, the history also provides several efficient iterators over the history of certain values. In general, these values are stored as `history.recent_NAME` and the iterator over them is just `history.NAME`. For example, `for addr in state.history.bbl_addrs: print hex(addr)` will print out a basic block address trace for the binary, while `state.history.recent_bbl_addrs` is the list of basic blocks executed in the most recent step, `state.history.parent.recent_bbl_addrs` is the list of basic blocks executed in the previous step, etc. If you ever need to quickly obtain a flat list of these values, you can access `.hardcopy`, e.g. `state.history.bbl_addrs.hardcopy`. Keep in mind though, index-based accessing is implemented on the iterators.

Here is a brief listing of some of the values stored in the history:

- `history.descriptions` is a listing of string descriptions of each of the rounds of execution performed on the state.
- `history.bbl_addrs` is a listing of the basic block addresses executed by the state. There may be more than one per round of execution, and not all addresses may correspond to binary code - some may be addresses at which SimProcedures are hooked.
- `history.jumpkinds` is a listing of the disposition of each of the control flow transitions in the state's history, as VEX enum strings.
- `history.jump_guards` is a listing of the conditions guarding each of the branches that the state has encountered.
- `history.events` is a semantic listing of "interesting events" which happened during execution, such as the presence of a symbolic jump condition, the program popping up a message box, or

execution terminating with an exit code.

- `history.actions` is usually empty, but if you add the `angr.options.refs` options to the state, it will be populated with a log of all the memory, register, and temporary value accesses performed by the program.

The callstack plugin

angr will track the call stack for the emulated program. On every call instruction, a frame will be added to the top of the tracked callstack, and whenever the stack pointer drops below the point where the topmost frame was called, a frame is popped. This allows angr to robustly store data local to the current emulated function.

Similar to the history, the callstack is also a linked list of nodes, but there are no provided iterators over the contents of the nodes - instead you can directly iterate over `state.callstack` to get the callstack frames for each of the active frames, in order from most recent to oldest. If you just want the topmost frame, this is `state.callstack`.

- `callstack.func_addr` is the address of the function currently being executed
- `callstack.call_site_addr` is the address of the basic block which called the current function
- `callstack.stack_ptr` is the value of the stack pointer from the beginning of the current function
- `callstack.ret_addr` is the location that the current function will return to if it returns

More about I/O: Files, file systems, and network sockets

Please refer to Working with File System, Sockets, and Pipes for a more complete and detailed documentation of how I/O is modeled in angr.

Copying and Merging

A state supports very fast copies, so that you can explore different possibilities:

```
>>> proj = angr.Project('/bin/true')
>>> s = proj.factory.blank_state()
>>> s1 = s.copy()
>>> s2 = s.copy()

>>> s1.mem[0x1000].uint32_t = 0x41414141
>>> s2.mem[0x1000].uint32_t = 0x42424242
```

States can also be merged together.

```
# merge will return a tuple. the first element is the merged state
# the second element is a symbolic variable describing a state flag
# the third element is a boolean describing whether any merging was done
>>> (s_merged, m, anything_merged) = s1.merge(s2)

# this is now an expression that can resolve to "AAAA" *or* "BBBB"
>>> aaaa_or_bbbb = s_merged.mem[0x1000].uint32_t
```

TODO: describe limitations of merging

Simulation Managers

The most important control interface in angr is the SimulationManager, which allows you to control symbolic execution over groups of states simultaneously, applying search strategies to explore a program's state space. Here, you'll learn how to use it.

Simulation managers let you wrangle multiple states in a slick way. States are organized into “stashes”, which you can step forward, filter, merge, and move around as you wish. This allows you to, for example, step two different stashes of states at different rates, then merge them together. The default stash for most operations is the `active` stash, which is where your states get put when you initialize a new simulation manager.

Stepping

The most basic capability of a simulation manager is to step forward all states in a given stash by one basic block. You do this with `.step()`.

```
>>> import angr
>>> proj = angr.Project('examples/fauxware/fauxware', auto_load_libs=False)
>>> state = proj.factory.entry_state()
>>> simgr = proj.factory.simgr(state)
>>> simgr.active
[<SimState @ 0x400580>]

>>> simgr.step()
>>> simgr.active
[<SimState @ 0x400540>]
```

Of course, the real power of the stash model is that when a state encounters a symbolic branch condition, both of the successor states appear in the stash, and you can step both of them in sync. When you don't really care about controlling analysis very carefully and you just want to step until there's nothing left to step, you can just use the `.run()` method.

```
# Step until the first symbolic branch
>>> while len(simgr.active) == 1:
...    simgr.step()

>>> simgr
<SimulationManager with 2 active>
>>> simgr.active
```

```
[<SimState @ 0x400692>, <SimState @ 0x400699>]

# Step until everything terminates
>>> simgr.run()
>>> simgr
<SimulationManager with 3 deadended>
```

We now have 3 deadended states! When a state fails to produce any successors during execution, for example, because it reached an `exit` syscall, it is removed from the active stash and placed in the `deadended` stash.

Stash Management

Let's see how to work with other stashes.

To move states between stashes, use `.move()`, which takes `from_stash`, `to_stash`, and `filter_func` (optional, default is to move everything). For example, let's move everything that has a certain string in its output:

```
>>> simgr.move(from_stash='deadended', to_stash='authenticated', filter_func=lambda s: b'W
>>> simgr
<SimulationManager with 2 authenticated, 1 deadended>
```

We were able to just create a new stash named "authenticated" just by asking for states to be moved to it. All the states in this stash have "Welcome" in their stdout, which is a fine metric for now.

Each stash is just a list, and you can index into or iterate over the list to access each of the individual states, but there are some alternate methods to access the states too. If you prepend the name of a stash with `one_`, you will be given the first state in the stash. If you prepend the name of a stash with `mp_`, you will be given a mulpyplexed version of the stash.

```
>>> for s in simgr.deadended + simgr.authenticated:
...     print(hex(s.addr))
0x1000030
0x1000078
0x1000078

>>> simgr.one_deadended
<SimState @ 0x1000030>
>>> simgr.mp_authenticated
MP([<SimState @ 0x1000078>, <SimState @ 0x1000078>])
>>> simgr.mp_authenticated.posix.dumps(0)
MP(['\x00\x00\x00\x00\x00\x00\x00\x00\x00SOSNEAKY\x00',
    '\x00\x00\x00\x00\x00\x00\x00\x00\x00S\x80\x80\x80\x80@\x80@\x00'])
```

Of course, `step`, `run`, and any other method that operates on a single stash of paths can take a `stash` argument, specifying which stash to operate on.

There are lots of fun tools that the simulation manager provides you for managing your stashes. We won't go into the rest of them for now, but you should check out the API documentation. TODO: link

Stash types

You can use stashes for whatever you like, but there are a few stashes that will be used to categorize some special kinds of states. These are:

Stash	Description
active	This stash contains the states that will be stepped by default, unless an alternate stash is specified.
deadended	A state goes to the deadended stash when it cannot continue the execution for some reason, including no more valid instructions, unsat state o all of its successors, or an invalid instruction pointer.
pruned	When using `LAZY_SOLVES`, states are not checked for satisfiability unless absolutely necessary. When a state is found to be unsat in th presence of `LAZY_SOLVES`, the state hierarchy traversed to identify when, in its history, it initially became unsat. All states that are descendants of that point (which will also be unsat, since a state cannot become un-unsat) are pruned and put in this stash.
unconstrained	If the `save_unconstrained` option is provide to the SimulationManager constructor, states that are determined to be unconstrained (i.e., with the instruction pointer controlled by user data or some other source of symbolic data) are placed here.
unsat	If the `save_unsat` option is provided to the SimulationManager constructor, states that are determined to be unsatisfiable (i.e., they have constraints that are contradictory, like the input having to be both "AAAA" and "BBBB" at the sam time) are placed here.

There is another list of states that is not a stash: `errored`. If, during execution, an error is raised, then the state will be wrapped in an `ErrorRecord` object, which contains the state and the error it raised, and then the record will be inserted into `errored`. You can get at the state as it was at the beginning of the execution tick that caused the error with `record.state`, you can see the error that was raised with `record.error`, and you can launch a debug shell at the site of the error with `record.debug()`. This is an invaluable debugging tool!

Simple Exploration

An extremely common operation in symbolic execution is to find a state that reaches a certain address, while discarding all states that go through another address. Simulation manager has a shortcut for this pattern, the `.explore()` method.

When launching `.explore()` with a `find` argument, execution will run until a state is found that matches the find condition, which can be the address of an instruction to stop at, a list of addresses to stop at, or a function which takes a state and returns whether it meets some criteria. When any of the states in the active stash match the `find` condition, they are placed in the `found` stash, and execution terminates. You can then explore the found state, or decide to discard it and continue with the other ones. You can also specify an `avoid` condition in the same format as `find`. When a state matches the avoid condition, it is put in the `avoided` stash, and execution continues. Finally, the `num_find` argument controls the number of states that should be found before returning, with a default of 1. Of course, if you run out of states in the active stash before finding this many solutions, execution will stop anyway.

Let's look at a simple crackme example:

First, we load the binary.

```
>>> proj = angr.Project('examples/CSCI-4968-MBE/challenges/crackme0x00a/crackme0x00a')
```

Next, we create a SimulationManager.

```
>>> simgr = proj.factory.simgr()
```

Now, we symbolically execute until we find a state that matches our condition (i.e., the "win" condition).

```
>>> simgr.explore(find=lambda s: b"Congrats" in s.posix.dumps(1))
<SimulationManager with 1 active, 1 found>
```

Now, we can get the flag out of that state!

```
>>> s = simgr.found[0]
>>> print(s.posix.dumps(1))
Enter password: Congrats!

>>> flag = s.posix.dumps(0)
>>> print(flag)
g00dJ0B!
```

Pretty simple, isn't it?

Other examples can be found by browsing the examples.

Exploration Techniques

angr ships with several pieces of canned functionality that let you customize the behavior of a simulation manager, called *exploration techniques*. The archetypical example of why you would want an exploration technique is to modify the pattern in which the state space of the program is explored - the default "step everything at once" strategy is effectively breadth-first search, but with an exploration technique you could implement, for example, depth-first search. However, the instrumentation power of these techniques is much more flexible than that - you can totally alter the behavior of angr's stepping process. Writing your own exploration techniques will be covered in a later chapter.

To use an exploration technique, call `simgr.use_technique(tech)`, where tech is an instance of an ExplorationTechnique subclass. angr's built-in exploration techniques can be found under `angr.exploration_techniques`.

Here's a quick overview of some of the built-in ones:

- *DFS*: Depth first search, as mentioned earlier. Keeps only one state active at once, putting the rest in the `deferred` stash until it deadends or errors.
- *Explorer*: This technique implements the `.explore()` functionality, allowing you to search for and avoid addresses.
- *LengthLimiter*: Puts a cap on the maximum length of the path a state goes through.
- *LoopSeer*: Uses a reasonable approximation of loop counting to discard states that appear to be going through a loop too many times, putting them in a `spinning` stash and pulling them out again if we run out of otherwise viable states.
- *ManualMergepoint*: Marks an address in the program as a merge point, so states that reach that address will be briefly held, and any other states that reach that same point within a timeout will be merged together.
- *MemoryWatcher*: Monitors how much memory is free/available on the system between simgr steps and stops exploration if it gets too low.
- *Oppologist*: The "operation apologist" is an especially fun gadget - if this technique is enabled and angr encounters an unsupported instruction, for example a bizzare and foreign floating point SIMD op, it will concretize all the inputs to that instruction and emulate the single instruction using the unicorn engine, allowing execution to continue.
- *Spiller*: When there are too many states active, this technique can dump some of them to disk in order to keep memory consumption low.
- *Threading*: Adds thread-level parallelism to the stepping process. This doesn't help much because of python's global interpreter locks, but if you have a program whose analysis spends a lot of time in angr's native-code dependencies (unicorn, z3, libvex) you can seem some gains.
- *Tracer*: An exploration technique that causes execution to follow a dynamic trace recorded from some other source. The dynamic tracer repository has some tools to generate those traces.
- *Veritesting*: An implementation of a CMU paper on automatically identifying useful merge points. This is

  so useful, you can enable it automatically with `veritesting=True` in the SimulationManager constructor! Note that it frequenly doesn't play nice with other techniques due to the invasive way it implements static symbolic execution.

Look at the API documentation for the simulation manager and exploration techniques for more information.

Execution Engines

When you ask for a step of execution to happen in angr, something has to actually perform the step. angr uses a series of engines (subclasses of the `SimEngine` class) to emulate the effects that of a given section of code has on an input state. The execution core of angr simply tries all the available engines in sequence, taking the first one that is able to handle the step. The following is the default list of engines, in order:

- The failure engine kicks in when the previous step took us to some uncontinuable state
- The syscall engine kicks in when the previous step ended in a syscall
- The hook engine kicks in when the current address is hooked
- The unicorn engine kicks in when the `UNICORN` state option is enabled and there is no symbolic data in the state
- The VEX engine kicks in as the final fallback.

SimSuccessors

The code that actually tries all the engines in turn is `project.factory.successors(state, **kwargs)`, which passes its arguments onto each of the engines. This function is at the heart of `state.step()` and `simulation_manager.step()`. It returns a SimSuccessors object, which we discussed briefly before. The purpose of SimSuccessors is to perform a simple categorization of the successor states, stored in various list attributes. They are:

Attribute	Guard Condition	Instruction Pointer	Description
`successors`	True (can be symbolic, but constrained to True)	Can be symbolic (but 256 solutions or less; see `unconstrained_successors`).	A normal, satisfiable successor state to the state processed by th engine. The instructic pointer of this state may be symbolic (i.e. a computed jump based on user input), so the state might actually represent *several* potential continuations of

			execution going
`unsat_successors`	False (can be symbolic, but constrained to False).	Can be symbolic.	Unsatisfiable successors. These ar successors whose guard conditions can only be false (i.e., jumps that cannot be taken, or the default branch of jumps that *must* be taken).
`flat_successors`	True (can be symbolic, but constrained to True).	Concrete value.	As noted above, state in the `successors` list can have symboli instruction pointers. This is rather confusing, as elsewhere in the code (i.e., in `SimEngineVEX.prcess`, when it's time step that state forward we make assumption that a single program state only represents the execution of a single spot in the cod To alleviate this, when we encounter states i `successors` with symbolic instruction pointers, we compute all possible concrete solutions (up to an arbitrary threshold of 256) for them, and make a copy of the state for each such solution. We call this process "flattening". These `flat_successors` are states, each of which has a different, concrete instruction pointer. For example, the instruction pointer of a state in `successors` was

			`X+5`, where `X` had constraints of `X > 0x800000` and `X < 0x800010`, we woul flatten it into 16 different `flat_successors` states, one with an instruction pointer of `0x800006`, one wit `0x800007`, and so on until `0x800015`.
`unconstrained_successors`	True (can be symbolic, but constrained to True).	Symbolic (with more than 256 solutions).	During the flattening procedure described above, if it turns out that there are more than 256 possible solutions for the instruction pointer, we assume that the instruction pointer has been overwritten with unconstrained data (i.e., a stack overflow with user data). *This assumption is not sound in general*. Su states are placed in `unconstrained_successors` and not i `successors`.
`all_successors`	Anything	Can be symbolic.	This is `successors` `+` `unsat_successor` `+` `unconstrained_successors`.

Breakpoints

TODO: rewrite this to fix the narrative

Like any decent execution engine, angr supports breakpoints. This is pretty cool! A point is set as follows:

```
>>> import angr
>>> b = angr.Project('examples/fauxware/fauxware')

# get our state
>>> s = b.factory.entry_state()

# add a breakpoint. This breakpoint will drop into ipdb right before a memory write happens
>>> s.inspect.b('mem_write')

# on the other hand, we can have a breakpoint trigger right *after* a memory write happens
# we can also have a callback function run instead of opening ipdb.
>>> def debug_func(state):
...     print("State %s is about to do a memory write!")

>>> s.inspect.b('mem_write', when=angr.BP_AFTER, action=debug_func)

# or, you can have it drop you in an embedded IPython!
>>> s.inspect.b('mem_write', when=angr.BP_AFTER, action=angr.BP_IPYTHON)
```

There are many other places to break than a memory write. Here is the list. You can break at BP_BEFORE or BP_AFTER for each of these events.

Event type	Event meaning
mem_read	Memory is being read.
mem_write	Memory is being written.
address_concretization	A symbolic memory access is being resolved.
reg_read	A register is being read.
reg_write	A register is being written.
tmp_read	A temp is being read.
tmp_write	A temp is being written.
expr	An expression is being created (i.e., a result of an arithmetic operation or a constant in the IR).
statement	An IR statement is being translated.
instruction	A new (native) instruction is being translated.
irsb	A new basic block is being translated.
constraints	New constraints are being added to the state.
exit	A successor is being generated from execution.

fork	A symbolic execution state has forked into multipl states.
symbolic_variable	A new symbolic variable is being created.
call	A call instruction is hit.
return	A ret instruction is hit.
simprocedure	A simprocedure (or syscall) is executed.
dirty	A dirty IR callback is executed.

These events expose different attributes:

Event type	Attribute name	Attribute availability	Attribute meaning
mem_read	mem_read_address	BP_BEFORE or BP_AFTER	The address at which memory is being reac
mem_read	mem_read_expr	BP_AFTER	The expression at tha address.
mem_read	mem_read_length	BP_BEFORE or BP_AFTER	The length of the memory read.
mem_read	mem_read_condition	BP_BEFORE or BP_AFTER	The condition of the memory read.
mem_write	mem_write_address	BP_BEFORE or BP_AFTER	The address at which memory is being written.
mem_write	mem_write_length	BP_BEFORE or BP_AFTER	The length of the memory write.
mem_write	mem_write_expr	BP_BEFORE or BP_AFTER	The expression that i being written.
mem_write	mem_write_condition	BP_BEFORE or BP_AFTER	The condition of the memory write.
reg_read	reg_read_offset	BP_BEFORE or BP_AFTER	The offset of the register being read.
reg_read	reg_read_length	BP_BEFORE or BP_AFTER	The length of the register read.
reg_read	reg_read_expr	BP_AFTER	The expression in the

			register
reg_read	reg_read_condition	BP_BEFORE or BP_AFTER	The condition of the register read.
reg_write	reg_write_offset	BP_BEFORE or BP_AFTER	The offset of the register being written.
reg_write	reg_write_length	BP_BEFORE or BP_AFTER	The length of the register write.
reg_write	reg_write_expr	BP_BEFORE or BP_AFTER	The expression that is being written.
reg_write	reg_write_condition	BP_BEFORE or BP_AFTER	The condition of the register write.
tmp_read	tmp_read_num	BP_BEFORE or BP_AFTER	The number of the temp being read.
tmp_read	tmp_read_expr	BP_AFTER	The expression of the temp.
tmp_write	tmp_write_num	BP_BEFORE or BP_AFTER	The number of the temp written.

These attributes can be accessed as members of `state.inspect` during the appropriate breakpoint callback to access the appropriate values. You can even modify these value to modify further uses of the values!

```
>>> def track_reads(state):
...     print('Read', state.inspect.mem_read_expr, 'from', state.inspect.mem_read_address)
...
>>> s.inspect.b('mem_read', when=angr.BP_AFTER, action=track_reads)
```

Additionally, each of these properties can be used as a keyword argument to `inspect.b` to make the breakpoint conditional:

```
# This will break before a memory write if 0x1000 is a possible value of its target express
>>> s.inspect.b('mem_write', mem_write_address=0x1000)

# This will break before a memory write if 0x1000 is the *only* value of its target express
>>> s.inspect.b('mem_write', mem_write_address=0x1000, mem_write_address_unique=True)

# This will break after instruction 0x8000, but only 0x1000 is a possible value of the las
>>> s.inspect.b('instruction', when=angr.BP_AFTER, instruction=0x8000, mem_read_expr=0x100
```

Cool stuff! In fact, we can even specify a function as a condition:

```
# this is a complex condition that could do anything! In this case, it makes sure that RAX
# that the basic block starting at 0x8004 was executed sometime in this path's history
>>> def cond(state):
...     return state.eval(state.regs.rax, cast_to=str) == 'AAAA' and 0x8004 in state.inspe

>>> s.inspect.b('mem_write', condition=cond)
```

That is some cool stuff!

Caution about `mem_read` breakpoint

The `mem_read` breakpoint gets triggered anytime there are memory reads by either the executing program or the binary analysis. If you are using breakpoint on `mem_read` and also using `state.mem` to load data from memory addresses, then know that the breakpoint will be fired as you are technically reading memory.

So if you want to load data from memory and not trigger any `mem_read` breakpoint you have had set up, then use `state.memory.load` with the keyword arguments `disable_actions=True` and `inspect=False`.

This is also true for `state.find` and you can use the same keyword arguments to prevent `mem_read` breakpoints from firing

Analyses

angr's goal is to make it easy to carry out useful analyses on binary programs. To this end, angr allows you to package analysis code in a common format that can be easily applied to any project. We will cover writing your own analyses later, but the idea is that all the analyses appear under `project.analyses` (for example, `project.analyses.CFGFast()`) and can be called as functions, returning analysis result instances.

Built-in Analyses

Name	Description
CFGFast	Constructs a fast *Control Flow Graph* of the program
CFGEmulated	Constructs an accurate *Control Flow Graph* of the program
VFG	Performs VSA on every function of the program, creating a *Value Flow Graph* and detecting stack variables

DDG	Calculates a *Data Dependency Graph*, allowing one to determine what statements a given value depends on
BackwardSlice	Computes a *Backward Slice* of a program with respect to a certain target
Identifier	Identifies common library functions in CGC binari
More!	angr has quite a few analyses, most of which worl If you'd like to know how to use one, please subm an issue requesting documentation.

Resilience

Analyses can be written to be resilient, and catch and log basically any error. These errors, depending on how they're caught, are logged to the `errors` or `named_errors` attribute of the analysis. However, you might want to run an analysis in "fail fast" mode, so that errors are not handled. To do this, the argument `fail_fast=True` can be passed into the analysis constructor.

Remarks

Congratulations! If you've read this far through the book (editor's note: this comment only really applies when we've actually finished writing all the TODOs so far) then you've been introduced to all the fundamental components of angr necessary to get started with binary analysis.

Ultimately, angr is just an emulator. It is a highly instrumentable and very unique emulator with lots of considerations for environment, true, but at its core, the work you do with angr is about extracting knowledge about how a bunch of bytecode behaves on a CPU. In designing angr, we've tried to provide you with the tools and abstractions on top of this emulator to make certain common tasks more useful, but there's no problem you can't solve just by working with a SimState and observing the affects of `.step()`.

As you read further into this book, we'll describe more technical subjects and how to tune angr's behavior for complicated scenarios. This knowledge should inform your use of angr so you can take the quickest path to a solution to any given problem, but ultimately, you will want to solve problems by exercising creativity with the tools at your disposal. If you can take a problem and wrangle it into a form where it has defined and tractable inputs and outputs, you can absolutely use angr to achieve your goals, given that these goals involve analyzing binaries. None of the abstractions or instrumentations we provide are the end-all of how to use angr for a given task - angr is designed so it can be used in as integrated or as ad-hoc of a manner as you desire. If you see a path from problem to solution, take it.

Of course, it's very difficult to become well-acquainted with such a huge piece of technology as angr. To this end you can absolutely lean on the community (through the angr slack is the best option) to discuss angr and solving problems with it.

Good luck!

Built-in Analyses

CFG

angr includes analyses to recover the control-flow graph of a binary program. This also includes recovery of function boundaries, as well as reasoning about indirect jumps and other useful metadata.

General ideas

A basic analysis that one might carry out on a binary is a Control Flow Graph. A CFG is a graph with (conceptually) basic blocks as nodes and jumps/calls/rets/etc as edges.

In angr, there are two types of CFG that can be generated: a static CFG (CFGFast) and a dynamic CFG (CFGEmulated).

CFGFast uses static analysis to generate a CFG. It is significantly faster, but is theoretically bounded by the fact that some control-flow transitions can only be resolved at execution-time. This is the same sort of CFG analysis performed by other popular reverse-engineering tools, and its results are comparable with their output.

CFGEmulated uses symbolic execution to capture the CFG. While it is theoretically more accurate, it is dramatically slower. It is also typically less complete, due to issues with the accuracy of emulation (system calls, missing hardware features, and so on)

If you are unsure which CFG to use, or are having problems with CFGEmulated, try CFGFast first.

A CFG can be constructed by doing:

```
>>> import angr
# load your project
>>> p = angr.Project('/bin/true', load_options={'auto_load_libs': False})

# Generate a static CFG
>>> cfg = p.analyses.CFGFast()

# generate a dynamic CFG
>>> cfg = p.analyses.CFGEmulated(keep_state=True)
```

Using the CFG

The CFG, at its core, is a NetworkX di-graph. This means that all of the normal NetworkX APIs are available:

```
>>> print("This is the graph:", cfg.graph)
>>> print("It has %d nodes and %d edges" % (len(cfg.graph.nodes()), len(cfg.graph.edges()))
```

The nodes of the CFG graph are instances of class `CFGNode`. Due to context sensitivity, a given basic block can have multiple nodes in the graph (for multiple contexts).

```
# this grabs *any* node at a given location:
>>> entry_node = cfg.get_any_node(p.entry)

# on the other hand, this grabs all of the nodes
>>> print("There were %d contexts for the entry block" % len(cfg.get_all_nodes(p.entry)))

# we can also look up predecessors and successors
>>> print("Predecessors of the entry point:", entry_node.predecessors)
>>> print("Successors of the entry point:", entry_node.successors)
>>> print("Successors (and type of jump) of the entry point:", [ jumpkind + " to " + str(n
```

Viewing the CFG

Control-flow graph rendering is a hard problem. angr does not provide any built-in mechanism for rendering the output of a CFG analysis, and attempting to use a traditional graph rendering library, like matplotlib, will result in an unusable image.

One solution for viewing angr CFGs is found in axt's angr-utils repository.

Shared Libraries

The CFG analysis does not distinguish between code from different binary objects. This means that by default, it will try to analyze control flow through loaded shared libraries. This is almost never intended behavior, since this will extend the analysis time to several days, probably. To load a binary without shared libraries, add the following keyword argument to the `Project` constructor: `load_options={'auto_load_libs': False}`

Function Manager

The CFG result produces an object called the *Function Manager*, accessible through `cfg.kb.functions`. The most common use case for this object is to access it like a dictionary. It maps addresses to `Function` objects, which can tell you properties about a function.

```
>>> entry_func = cfg.kb.functions[p.entry]
```

Functions have several important properties!

- `entry_func.block_addrs` is a set of addresses at which basic blocks belonging to the function begin.
- `entry_func.blocks` is the set of basic blocks belonging to the function, that you can explore and disassemble using capstone.
- `entry_func.string_references()` returns a list of all the constant strings that were referred to at any point in the function.
  They are formatted as `(addr, string)` tuples, where addr is the address in the binary's data section the string lives, and string is a python string that contains the value of the string.
- `entry_func.returning` is a boolean value signifying whether or not the function can return.
  `False` indicates that all paths do not return.
- `entry_func.callable` is an angr Callable object referring to this function.
  You can call it like a python function with python arguments and get back an actual result (may be symbolic) as if you ran the function with those arguments!
- `entry_func.transition_graph` is a NetworkX DiGraph describing control flow within the function itself. It resembles the control-flow graphs IDA displays on a per-function level.
- `entry_func.name` is the name of the function.
- `entry_func.has_unresolved_calls` and `entry.has_unresolved_jumps` have to do with detecting imprecision within the CFG.
  Sometimes, the analysis cannot detect what the possible target of an indirect call or jump could be.
  If this occurs within a function, that function will have the appropriate `has_unresolved_*` value set to `True`.
- `entry_func.get_call_sites()` returns a list of all the addresses of basic blocks which end in calls out to other functions.
- `entry_func.get_call_target(callsite_addr)` will, given `callsite_addr` from the list of call site addresses, return where that callsite will call out to.
- `entry_func.get_call_return(callsite_addr)` will, given `callsite_addr` from the list of call site addresses, return where that callsite should return to.

and many more !

CFGFast details

CFGFast peforms a static control-flow and function recovery. Starting with the entry point (or any user-defined points) roughly the following procedure is performed:

1) The basic block is lifted to VEX IR, and all its exits (jumps, calls, returns, or continuation to the next block) are collected 2) For each exit, if this exit is a constant address, we add an edge to the CFG of the correct type, and add the destination block to the set of blocks to be analyzed. 3) In the event of a function call, the destination block is also considered the start of a new function. If the target function is known to return, the block after the call is also analyzed. 4) In the event of a return, the current function is marked as returning,

and the appropriate edges in the callgraph and CFG are updated. 4) For all indirect jumps (block exits with a non-constant destination) Indirect Jump Resolution is performed.

Finding function starts

CFGFast supports multiple ways of deciding where a function starts and ends.

First the binary's main entry point will be analyzed. For binaries with symbols (e.g., non-stripped ELF and PE binaries) all function symbols will be used as possible starting points. For binaries without symbols, such as stripped binaries, or binaries loaded using the `blob` loader backend, CFG will scan the binary for a set of function prologues defined for the binary's architecture. Finally, by default, the binary's entire code section will be scanned for executable contents, regardless of prologues or symbols.

In addition to these, as with CFGEmulated, function starts will also be considered when they are the target of a "call" instruction on the given architecture.

All of these options can be disabled

FakeRets and function returns

When a function call is observed, we first assume that the callee function eventually returns, and treat the block after it as part of the caller function. This inferred control-flow edge is known as a "FakeRet". If, in analyzing the callee, we find this not to be true, we update the CFG, removing this "FakeRet", and updating the callgraph and function blocks accordingly. As such, the CFG is recovered *twice*. In doing this, the set of blocks in each function, and whether the function returns, can be recovered and propagated directly.

Indirect Jump Resolution

TODO

Options

These are the most useful options when working with CFGFast:

Option	Description
force_complete_scan	(Default: True) Treat the entire binary as code for the purposes of function detection. If you have a blob (e.g., mixed code and data) *you want to turn this off*.
function_starts	A list of addresses, to use as entry points into the analysis.
normalize	(Default: False) Normalize the resulting functions (e.g., each basic block belongs to at most one function, back-edges point to the start of basic blocks)
	(Default: True) Perform additional analysis to

resolve_indirect_jumps	attempt to find targets for every indirect jump foun during CFG creation
more!	Examine the docstring on p.analyses.CFGFast fo more up-to-date options

CFGEmulated details

Options

The most common options for CFGEmulated include:

Option	Description
context_sensitivity_level	This sets the context sensitivity level of the analysis. See the context sensitivity level section below for more information. This is 1 by default.
starts	A list of addresses, to use as entry points into the analysis.
avoid_runs	A list of addresses to ignore in the analysis.
call_depth	Limit the depth of the analysis to some number calls. This is useful for checking which functions a specific function can directly jump to (by setting `call_depth` to 1).
initial_state	An initial state can be provided to the CFG, which will use throughout its analysis.
keep_state	To save memory, the state at each basic block is discarded by default. If `keep_state` is True, the state is saved in the CFGNode.
enable_symbolic_back_traversal	Whether to enable an intensive technique for resolving indirect jumps
enable_advanced_backward_slicing	Whether to enable another intensive technique fo resolving direct jumps
	Examine the docstring on

more!	p.analyses.CFGEmulated for more up-to-date options

Context Sensitivity Level

angr constructs a CFG by executing every basic block and seeing where it goes. This introduces some challenges: a basic block can act differently in different *contexts*. For example, if a block ends in a function return, the target of that return will be different, depending on different callers of the function containing that basic block.

The context sensitivity level is, conceptually, the number of such callers to keep on the callstack. To explain this concept, let's look at the following code:

```c
void error(char *error)
{
    puts(error);
}

void alpha()
{
    puts("alpha");
    error("alpha!");
}

void beta()
{
    puts("beta");
    error("beta!");
}

void main()
{
    alpha();
    beta();
}
```

The above sample has four call chains: `main>alpha>puts`, `main>alpha>error>puts` and `main>beta>puts`, and `main>beta>error>puts`. While, in this case, angr can probably execute both call chains, this becomes unfeasible for larger binaries. Thus, angr executes the blocks with states limited by the context sensitivity level. That is, each function is re-analyzed for each unique context that it is called in.

For example, the `puts()` function above will be analyzed with the following contexts, given different context sensitivity levels:

Level	Meaning	Contexts
0	Callee-only	`puts`

1	One caller, plus callee	`alpha>puts` `beta>puts` `error>puts`
2	Two callers, plus callee	`alpha>error>puts` `main>alpha>puts` `beta>error>puts` `main>beta>puts`
3	Three callers, plus callee	`main>alpha>error>puts` `main>alpha>puts` `main>beta>error>puts` `main>beta>puts`

The upside of increasing the context sensitivity level is that more information can be gleaned from the CFG. For example, with context sensitivity of 1, the CFG will show that, when called from `alpha`, `puts` returns to `alpha`, when called from `error`, `puts` returns to `error`, and so forth. With context sensitivity of 0, the CFG simply shows that `puts` returns to `alpha`, `beta`, and `error`. This, specifically, is the context sensitivity level used in IDA. The downside of increasing the context sensitivity level is that it exponentially increases the analysis time.

Backward Slicing

A *program slice* is a subset of statements that is obtained from the original program, usually by removing zero or more statements. Slicing is often helpful in debugging and program understanding. For instance, it's usually easier to locate the source of a variable on a program slice.

A backward slice is constructed from a *target* in the program, and all data flows in this slice end at the *target*.

angr has a built-in analysis, called `BackwardSlice`, to construct a backward program slice. This section will act as a how-to for angr's `BackwardSlice` analysis, and followed by some in-depth discussion over the implementation choices and limitations.

First Step First

To build a `BackwardSlice`, you will need the following information as input.

- **Required** CFG. A control flow graph (CFG) of the program. This CFG must be an accurate CFG (CFGEmulated).
- **Required** Target, which is the final destination that your backward slice terminates at.
- **Optional** CDG. A control dependence graph (CDG) derived from the CFG.
  angr has a built-in analysis `CDG` for that purpose.
- **Optional** DDG. A data dependence graph (DDG) built on top of the CFG.
  angr has a built-in analysis `DDG` for that purpose.

A `BackwardSlice` can be constructed with the following code:

```
>>> import angr
# Load the project
>>> b = angr.Project("examples/fauxware/fauxware", load_options={"auto_load_libs": False})

# Generate a CFG first. In order to generate data dependence graph afterwards, you’ll have
# - keep all input states by specifying keep_state=True.
# - store memory, register and temporary values accesses by adding the angr.options.refs o
# Feel free to provide more parameters (for example, context_sensitivity_level) for CFG
# recovery based on your needs.
>>> cfg = b.analyses.CFGEmulated(keep_state=True,
...                              state_add_options=angr.sim_options.refs,
...                              context_sensitivity_level=2)

# Generate the control dependence graph
>>> cdg = b.analyses.CDG(cfg)

# Build the data dependence graph. It might take a while. Be patient!
>>> ddg = b.analyses.DDG(cfg)

# See where we wanna go... let’s go to the exit() call, which is modeled as a

# SimProcedure.
>>> target_func = cfg.kb.functions.function(name="exit")
# We need the CFGNode instance
>>> target_node = cfg.get_any_node(target_func.addr)

# Let’s get a BackwardSlice out of them!
# `targets` is a list of objects, where each one is either a CodeLocation
# object, or a tuple of CFGNode instance and a statement ID. Setting statement
# ID to -1 means the very beginning of that CFGNode. A SimProcedure does not
# have any statement, so you should always specify -1 for it.
>>> bs = b.analyses.BackwardSlice(cfg, cdg=cdg, ddg=ddg, targets=[ (target_node, -1) ])

# Here is our awesome program slice!
>>> print(bs)
```

Sometimes it’s difficult to get a data dependence graph, or you may simply want build a program slice on top of a CFG. That’s basically why DDG is an optional parameter. You can build a `BackwardSlice` solely based on CFG by doing:

```
>>> bs = b.analyses.BackwardSlice(cfg, control_flow_slice=True)
BackwardSlice (to [(<CFGNode exit (0x10000a0) [0]>, -1)])
```

Using The `BackwardSlice` Object

Before you go ahead and use `BackwardSlice` object, you should notice that the design of this class is fairly arbitrary right now, and it is still subject to change in the near future. We'll try our best to keep this documentation up-to-date.

Members

After construction, a `BackwardSlice` has the following members which describe a program slice:

Member	Mode	Meaning
runs_in_slice	CFG-only	A `networkx.DiGraph` instance showing addresses o blocks and SimProcedures in the program slice, as well as transitions between them
cfg_nodes_in_slice	CFG-only	A `networkx.DiGraph` instance showing CFGNodes i the program slice and transitio in between
chosen_statements	With DDG	A dict mapping basic block addresses to lists of statement IDs that are part of the program slice
chosen_exits	With DDG	A dict mapping basic block addresses to a list of "exits". Each exit in the list is a valid transition in the program slice

Each "exit" in `chosen_exit` is a tuple including a statement ID and a list of target addresses. For example, an "exit" might look like the following:

```
(35, [ 0x400020 ])
```

If the "exit" is the default exit of a basic block, it'll look like the following:

```
("default", [ 0x400085 ])
```

Export an Annotated Control Flow Graph

TODO

User-friendly Representation

Take a look at `BackwardSlice.dbg_repr()` !

TODO

Implementation Choices

TODO

Limitations

TODO

Completeness

TODO

Soundness

TODO

Function Identifier

The identifier uses test cases to identify common library functions in CGC binaries. It prefilters by finding some basic information about stack variables/arguments. The information of about stack variables can be generally useful in other projects.

```
>>> import angr

# get all the matches
>>> p = angr.Project("../binaries/tests/i386/identifiable")
# note analysis is executed via the Identifier call
>>> idfer = p.analyses.Identifier()
>>> for funcInfo in idfer.func_info:
...     print(hex(funcInfo.addr), funcInfo.name)

0x8048e60 memcmp
0x8048ef0 memcpy
0x8048f60 memmove
0x8049030 memset
0x8049320 fdprintf
0x8049a70 sprintf
0x8049f40 strcasecmp
```

```
0x804a0f0 strcmp
0x804a190 strcpy
0x804a260 strlen
0x804a3d0 strncmp
0x804a620 strtol
0x804aa00 strtol
0x80485b0 free
0x804aab0 free
0x804aad0 free
0x8048660 malloc
0x80485b0 free
```

Advanced Topics

Gotchas

This section contains a list of gotchas that users/victims of angr frequently run into.

SimProcedure inaccuracy

To make symbolic execution more tractable, angr replaces common library functions with summaries written in Python. We call these summaries SimProcedures. SimProcedures allow us to mitigate path explosion that would otherwise be introduced by, for example, `strlen` running on a symbolic string.

Unfortunately, our SimProcedures are far from perfect. If angr is displaying unexpected behavior, it might be caused by a buggy/incomplete SimProcedure. There are several things that you can do:

1. Disable the SimProcedure (you can exclude specific SimProcedures by passing options to the angr.Project class). This has the drawback of likely leading to a path explosion, unless you are very careful about constraining the input to the function in question. The path explosion can be partially mitigated with other angr capabilities (such as Veritesting).
2. Replace the SimProcedure with something written directly to the situation in question. For example, our `scanf` implementation is not complete, but if you just need to support a single, known format string, you can write a hook to do exactly that.
3. Fix the SimProcedure.

Unsupported syscalls

System calls are also implemented as SimProcedures. Unfortunately, there are system calls that we have not yet implemented in angr. There are several workarounds for an unsupported system call:

1. Implement the system call. *TODO: document this process*
2. Hook the callsite of the system call (using `project.hook`) to make the required modifications to the state in an ad-hoc way.
3. Use the `state.posix.queued_syscall_returns` list to queue syscall return values. If a return value is queued, the system call will not be executed, and the value will be used instead. Furthermore, a function can be queued instead as the "return value", which will result in that function being applied to the state when the system call is triggered.

Symbolic memory model

The default memory model used by angr is inspired by Mayhem. This memory model supports limited symbolic reads and writes. If the memory index of a read is symbolic and the range of possible values of this index is too wide, the index is concretized to a single value. If the memory index of a write is symbolic at all, the index is concretized to a single value. This is configurable by changing the memory concretization strategies of `state.memory`.

Symbolic lengths

SimProcedures, and especially system calls such as `read()` and `write()` might run into a situation where the *length* of a buffer is symbolic. In general, this is handled very poorly: in many cases, this length will end up being concretized outright or retroactively concretized in later steps of execution. Even in cases when it is not, the source or destination file might end up looking a bit "weird".

The Whole Pipeline

Understanding the Execution Pipeline

If you've made it this far you know that at its core, angr is a highly flexible and intensely instrumentable emulator. In order to get the most mileage out of it, you'll want to know what happens at every step of the way when you say `simgr.run()`.

This is intended to be a more advanced document; you'll need to understand the function and intent of `SimulationManager`, `ExplorationTechnique`, `SimState`, and `SimEngine` in order to understand what we're talking about at times! You may want to have the angr source open to follow along with this.

At every step along the way, each function will take `**kwargs` and pass them along to the next function in the hierarchy, so you can pass parameters to any point in the hierarchy and they will trickle down to everything below.

Simulation Managers

So you've set your analysis in motion. Time to begin our journey.

`run()`

`SimulationManager.run()` takes several optional parameters, all of which control when to break out of the stepping loop. Notably, `n`, and `until`. `n` is used immediately - the run function loops, calling the `step()` function and passing on all its parameters until either `n` steps have happened or some other termination condition has occurred. If `n` is not provided, it defaults to 1, unless an `until` function is provided, in which case there will be no numerical cap on the loop. Additionally, the stash that is being used is taken into consideration, as if it becomes empty execution must terminate.

So, in summary, when you call `run()`, `step()` will be called in a loop until any of the following:

1. The `n` number of steps have elapsed
2. The `until` function returns true
3. The exploration techniques `complete()` hooks (combined via the `SimulationManager.completion_mode` parameter/attribute - it is by default the `any` builtin function but can be changed to `all` for example) indicate that the analysis is complete
4. The stash being executed becomes empty

An aside: explore()

`SimulationManager.explore()` is a very thin wrapper around `run()` which adds the `Explorer` exploration technique, since performing one-off explorations is a very common action. Its code in its entirety is below:

```
num_find += len(self._stashes[find_stash]) if find_stash in self._stashes else 0
tech = self.use_technique(Explorer(find, avoid, find_stash, avoid_stash, cfg, num_find))

try:
    self.run(stash=stash, n=n, **kwargs)
finally:
    self.remove_technique(tech)

return self
```

Exploration technique hooking

From here down, every function in the simulation manager can be instrumented by an exploration technique. The exact mechanism through which this works is that when you call `SimulationManager.use_technique()`, angr monkeypatches the simulation manager to replace any function implemented in the exploration technique's body with a function which will first call the exploration technique's function, and then on the second call will call the original function. This is somewhat messy to implement and certainly not thread safe by any means, but does produce a clean and powerful interface for exploration techniques to instrument stepping behavior, either before or after the original function is called, even choosing whether or not to call the original function whatsoever. Additionally, it allows multiple exploration techniques to hook the same function, as the monkeypatched function simply becomes the "original" function for the next applied hook.

`step()`

There is a lot of complicated logic in `step()` to handle degenerate cases - mostly implementing the population of the `deadended` stash, the `save_unsat` option, and calling the `filter()` exploration technique hooks. Beyond this, though, most of the logic is looping through the stash specified by the `stash` argument and calling `step_state()` on each state, then applying the dict result of `step_state()` to the stash list. Finally, if the `step_func` parameter is provided, it is called with the simulation manager as a parameter before the step ends.

`step_state()`

The default `step_state()`, which can be overridden or instrumented by exploration techniques, is also simple - it calls `successors()`, which returns a `SimSuccessors` object, and then translates it into a dict mapping stash names to new states which should be added to that stash. It also implements error handling - if `successors()` throws an error, it will be caught and an `ErrorRecord` will be inserted into `SimulationManager.errored`.

`successors()`

We've almost made it out of SimulationManager. `successors()`, which can also be instrumented by exploration techniques, is supposed to take a state and step it forward, returning a `SimSuccessors` object categorizing its successors independently of any stash logic. If the `successor_func` parameter was provided, it is used and its return value is returned directly. If this parameter was not provided, we use the `project.factory.successors` method to tick the state forward and get our `SimSuccessors`.

The Engine

When we get to the actual successors generation, we need to figure out how to actually perform the execution. Hopefully, the angr documentation has been organized in a way such that by the time you reach this page, you know that a `SimEngine` is a device that knows how to take a state and produce its successors. There is only one "default engine" per project, but you can provide the `engine` parameter to specify which engine will be used to perform the step.

Keep in mind that this parameter can be provided way at the top, to `.step()`, `.explore()`, `.run()` or anything else that starts execution, and they will be filtered down to this level. Any additional parameters will continue being passed down, until they reach the part of the engine they are intended for. The engine will discard any parameters it doesn't understand.

Generally, the main entry point of an engine is `SimEngine.process()`, which can return whatever result it likes, but for simulation managers, engines are required to use `SuccessorsMixin`, which provides a `process()` method, which creates a `SimSuccessors` object and then calls `process_successors()` so that other mixins can fill it out.

angr's default engine, the `UberEngine`, contains several mixins which provide the `process_successors()` method:

- `SimEngineFailure` - handles stepping states with degenerate jumpkinds
- `SimEngineSyscall` - handles stepping states which have performed a syscall and need it executed
- `HooksMixin` - handles stepping states which have reached a hooked address and need the hook executed
- `SimEngineUnicorn` - executes machine code via the unicorn engine
- `SootMixin` - executes java bytecode via the SOOT IR
- `HeavyVEXMixin` - executes machine code via the VEX IR

Each of these mixins is implemented to fill out the `SimSuccessors` object if they can handle the current state, otherwise they call `super()` to pass the job on to the next class in the stack.

Engine mixins

`SimEngineFailure` handles error cases. It is only used when the previous jumpkind is one of `Ijk_EmFail`, `Ijk_MapFail`, `Ijk_Sig*`, `Ijk_NoDecode` (but only if the address is not hooked), or `Ijk_Exit`. In the first four cases, its action is to raise an exception. In the last case, its action is to simply produce no successors.

`SimEngineSyscall` services syscalls. It is used when the previous jumpkind is anything of the form `Ijk_Sys*`. It works by making a call into `SimOS` to retrieve the SimProcedure that should be run to respond to this syscall, and then running it! Pretty simple.

`HooksMixin` provides the hooking functionality in angr. It is used when a state is at an address that is hooked, and the previous jumpkind is *not* `Ijk_NoHook`. It simply looks up the associated SimProcedure and runs it on the state! It also takes the parameter `procedure`, which will cause the given procedure to be run for the current step even if the address is not hooked.

`SimEngineUnicorn` performs concrete execution with the Unicorn Engine. It is used when the state option `o.UNICORN` is enabled, and a myriad of other conditions designed for maximum efficiency (described below) are met.

`SootMixin` performs execution over the SOOT IR. Not very important unless you are analyzing java bytecode, in which case it is very important.

`SimEngineVEX` is the big fellow. It is used whenever any of the previous can't be used. It attempts to lift bytes from the current address into an IRSB, and then executes that IRSB symbolically. There are a huge number of parameters that can control this process, so I will merely link to the API reference describing them.

The exact process by which SimEngineVEX digs into an IRSB is a little complicated, but essentially it runs all the block's statements in order. This code is worth reading if you want to see the true inner core of angr's symbolic execution.

When using Unicorn Engine

If you add the `o.UNICORN` state option, at every step `SimEngineUnicorn` will be invoked, and try to see if it is allowed to use Unicorn to execute concretely.

What you REALLY want to do is to add the predefined set `o.unicorn` (lowercase) of options to your state:

```
unicorn = { UNICORN, UNICORN_SYM_REGS_SUPPORT, INITIALIZE_ZERO_REGISTERS, UNICORN_HANDLE_T
```

These will enable some additional functionalities and defaults which will greatly enhance your experience. Additionally, there are a lot of options you can tune on the `state.unicorn` plugin.

A good way to understand how unicorn works is by examining the logging output (`logging.getLogger('angr.engines.unicorn_engine').setLevel('DEBUG'); logging.getLogger('angr.state_plugins.unicorn_engine').setLevel('DEBUG')` from a sample run of unicorn.

```
INFO    | 2017-02-25 08:19:48,012 | angr.state_plugins.unicorn | started emulation at 0x40
```

Here, angr diverts to unicorn engine, beginning with the basic block at 0x4012f9. The maximum step count is set to 1000000, so if execution stays in Unicorn for 1000000 blocks, it'll automatically pop out. This is to avoid hanging in an infinite loop. The block count is configurable via the `state.unicorn.max_steps` variable.

```
INFO    | 2017-02-25 08:19:48,014 | angr.state_plugins.unicorn | mmap [0x401000, 0x401fff]
INFO    | 2017-02-25 08:19:48,016 | angr.state_plugins.unicorn | mmap [0x7fffffffffe0000, 
INFO    | 2017-02-25 08:19:48,019 | angr.state_plugins.unicorn | mmap [0x6010000, 0x601fff
INFO    | 2017-02-25 08:19:48,022 | angr.state_plugins.unicorn | mmap [0x602000, 0x602fff]
INFO    | 2017-02-25 08:19:48,023 | angr.state_plugins.unicorn | mmap [0x400000, 0x400fff]
INFO    | 2017-02-25 08:19:48,025 | angr.state_plugins.unicorn | mmap [0x7000000, 0x7000ff
```

angr performs lazy mapping of data that is accessed by unicorn engine, as it is accessed. 0x401000 is the page of instructions that it is executing, 0x7fffffffffe0000 is the stack, and so on. Some of these pages are symbolic, meaning that they contain at least some data that, when accessed, will cause execution to abort out of Unicorn.

```
INFO    | 2017-02-25 08:19:48,037 | angr.state_plugins.unicorn | finished emulation at 0x7
```

Execution stays in Unicorn for 3 basic blocks (a computational waste, considering the required setup), after which it reaches a simprocedure location and jumps out to execute the simproc in angr.

```
INFO    | 2017-02-25 08:19:48,076 | angr.state_plugins.unicorn | started emulation at 0x40
INFO    | 2017-02-25 08:19:48,077 | angr.state_plugins.unicorn | mmap [0x401000, 0x401fff]
INFO    | 2017-02-25 08:19:48,079 | angr.state_plugins.unicorn | mmap [0x7fffffffffe0000, 
INFO    | 2017-02-25 08:19:48,081 | angr.state_plugins.unicorn | mmap [0x6010000, 0x601fff
```

After the simprocedure, execution jumps back into Unicorn.

```
WARNING | 2017-02-25 08:19:48,082 | angr.state_plugins.unicorn | fetching empty page [0x0,
INFO    | 2017-02-25 08:19:48,103 | angr.state_plugins.unicorn | finished emulation at 0x4
```

Execution bounces out of Unicorn almost right away because the binary accessed the zero-page.

```
INFO    | 2017-02-25 08:19:48,120 | angr.engines.unicorn_engine | not enough runs since la
INFO    | 2017-02-25 08:19:48,125 | angr.engines.unicorn_engine | not enough runs since la
```

To avoid thrashing in and out of Unicorn (which is expensive), we have cooldowns (attributes of the `state.unicorn` plugin) that wait for certain conditions to hold (i.e., no symbolic memory accesses for X blocks) before jumping back into unicorn when a unicorn run is aborted due to anything but a simprocedure or syscall. Here, the condition it's waiting for is for 100 blocks to be executed before jumping back in.

The Mixin Pattern

If you are trying to work more intently with the deeper parts of angr, you will need to understand one of the design patterns we use frequently: the mixin pattern.

In brief, the mixin pattern is where python's subclassing features is used not to implement IS-A relationships (a Child is a kind of Person) but instead to implement pieces of functionality for a type in different classes to make more modular and maintainable code. Here's an example of the mixin pattern in action:

```python
class Base:
    def add_one(self, v):
        return v + 1

class StringsMixin(Base):
    def add_one(self, v):
        coerce = type(v) is str
        if coerce:
            v = int(v)
        result = super().add_one(v)
        if coerce:
            result = str(result)
        return result

class ArraysMixin(Base):
    def add_one(self, v):
        if type(v) is list:
            return [super().add_one(v_x) for v_x in v]
        else:
            return super().add_one(v)

class FinalClass(ArraysMixin, StringsMixin, Base):
    pass
```

With this construction, we are able to define a very simple interface in the `Base` class, and by "mixing in" two mixins, we can create the `FinalClass` which has the same interface but with additional features. This is accomplished through python's powerful multiple inheritance model, which handles method dispatch by creating a *method resolution order*, or MRO, which is unsuprisingly a list which determines the order in which methods are called as execution proceeds through `super()` calls. You can view a class' MRO as such:

```
FinalClass.__mro__

(FinalClass, ArraysMixin, StringsMixin, Base, object)
```

This means that when we take an instance of `FinalClass` and call `add_one()`, python first checks to see if `FinalClass` defines an `add_one`, and then `ArraysMixin`, and so on and so forth. Furthermore, when `ArraysMixin` calls `super().add_one()`, python will skip past `ArraysMixin` in the MRO, first checking if `StringsMixin` defines an `add_one`, and so forth.

Because multiple inheritance can create strange dependency graphs in the subclass relationship, there are rules for generating the MRO and for determining if a given mix of mixins is even allowed. This is important to understand when building complex classes with many mixins which have dependencies on each other. In short: left-to-right, depth-first, but deferring any base classes which are shared by multiple subclasses (the merge point of a diamond pattern in the inheritance graph) until the last point where they would be encountered in this depth-first search. For example, if you have classes A, B(A), C(B), D(A), E(C, D), then the method resolution order will be E, C, B, D, A. If there is any case in which the MRO would be ambiguous, the class construction is illegal and will throw an exception at import time.

This is complicated! If you find yourself confused, the canonical document explaining the rationale, history, and mechanics of python's multiple inheritence can be found here.

Mixins in Claripy Solvers

yan please write something here

Mixins in angr Engines

The main entry point to a SimEngine is `process()`, but how do we determine what that does?

The mixin model is used in SimEngine and friends in order to allow pieces of functionality to be reused between static and symbolic analyses. The default engine, `UberEngine`, is defined as follows:

```
class UberEngine(SimEngineFailure, SimEngineSyscall, HooksMixin, SimEngineUnicorn, SuperFa
    pass
```

Each of these mixins provides either execution through a different medium or some additional instrumentation feature. Though they are not listed here explicitly, there are some base classes implicit to this hierarchy which set up the way this class is traversed. Most of these mixins inherit from `SuccessorsMixin`, which is what provides the basic `process()` implementation. This function sets up the `SimSuccessors` for the rest of the mixins to fill in, and then calls `process_successors()`, which each of the mixins which provide some mode of execution implement. If the mixin can handle the step, it does so and returns, otherwise it calls `super().process_successors()`. In this way, the MRO for the engine class determines what the order of precedence for the engine's pieces is.

HeavyVEXMixin and friends

Let's take a closer look at the last mixin, `HeavyVEXMixin`. If you look at the module hierarchy of the angr `engines` submodule, you will see that the `vex` submodule has a lot of pieces in it which are organized by how tightly tied to particular state types or data types they are. The heavy VEX mixin is one version of the culmination of all of these. Let's look at its definition:

```
class HeavyVEXMixin(SuccessorsMixin, ClaripyDataMixin, SimStateStorageMixin, VEXMixin, VEX
    ...
    # a WHOLE lot of implementation
```

So, the heavy VEX mixin is meant to provide fully instrumented symbolic execution on a SimState. What does this entail? The mixins tell the tale.

First, the plain `VEXMixin`. This mixin is designed to provide the barest-bones framework for processing a VEX block. Take a look at its source code. Its main purpose is to perform the preliminary digestion of the VEX IRSB and dispatch processing of it to methods which are provided by mixins - look at the methods which are either `pass` or `return NotImplemented`. Notice that absolutely none of its code makes any assumption whatsoever of what the type of `state` is or even what the type of the data words inside `state` are. This job is delegated to other mixins, making the `VEXMixin` an appropriate base class for literally any analysis on VEX blocks.

The next-most interesting mixin is the `ClaripyDataMixin`, whose source code is here. This mixin actually integrates the fact that we are executing over the domain of Claripy ASTs. It does this by implementing some of the methods which are unimplemented in the `VEXMixin`, most importantly the `ITE` expression, all the operations, and the clean helpers.

In terms of what it looks like to actually touch the SimState, the `SimStateStorageMixin` provides the glue between the `VEXMixin`'s interface for memory writes et al and SimState's interface for memory writes and such. It is unremarkable, except for a small interaction between it and the `ClaripyDataMixin`. The Claripy mixin also overrides the memory/register read/write functions, for the purpose of converting between the bitvector and floating-point types, since the vex interface expects to be able to load and store floats, but the SimState interface wants to load and store only bitvectors. Because of this, *the claripy mixin must come before the storage mixin in the MRO*. This is very much an interaction like the one in the add_one example at the start of this page - one mixin serves as a data filtering layer for another mixin.

Instrumenting the data layer

Let's turn our attention to a mixin which is not included in the `HeavyVEXMixin` but rather mixed into the `UberEngine` formula explicitly: the `TrackActionsMixin`. This mixin implements "SimActions", which is angr parlance for dataflow tracking. Again, look at the source code. The way it does this is that it *wraps and unwraps the data layer* to pass around additional information about data flows. Look at how it instruments `RdTmp`, for instance. It immediately `super()`-calls to the next method in the MRO, but instead of returning that data it returns a tuple of the data and its dependencies, which depending on whether you want temporary variables to be atoms in the dataflow model, will either be just the tmp which was read or the dependencies of the value written to that tmp.

This pattern continues for every single method that this mixin touches - any expression it receives must be unpacked into the expression and its dependencies, and any result must be packaged with its dependencies before it is returned. This works because the mixin above it makes no assumptions about what data it is passing around, and the mixin below it never gets to see any dependencies whatsoever. In fact, there could be multiple mixins performing this kind of wrap-unwrap trick and they could all coexist peacefully!

Note that a mixin which instruments the data layer in this way is *obligated* to override *every single method which takes or returns an expression value*, even if it doesn't perform any operation on the expression other than doing the wrapping and unwrapping. To understand why, imagine that the mixin does not override the `_handle_vex_const` expression, so immediate value loads are not annotated with dependencies. The expression value which will be returned from the mixin which does provide `_handle_vex_const` will not be a tuple of (expression, deps), it will just be the expression. Imagine this execution is taking place in the context of a `WrTmp(t0, Const(0))`. The const expression will be passed down to the `WrTmp` handler along with the identifier of the tmp to write to. However, since `_handle_vex_stmt_WrTmp` *will* be overridden by our mixin which touches the data layer, it expects to be passed the tuple including the deps, and so it will crash when trying to unpack the not-a-tuple value.

In this way, you can sort of imagine that a mixin which instruments the data layer in this way is actually creating a contract within python's nonexistent typesystem - you are guaranteed to receive back any types you return, but you must pass down any types you receive as return values from below.

Mixins in the memory model

audrey please write something here. or fish, I'm not picky

Optimizing Symbolic Execution

The performance of angr as an analysis tool or emulator is greatly handicapped by the fact that lots of it is written in python. Regardless, there are a lot of optimizations and tweaks you can use to make angr faster and lighter.

General speed tips

- *Use pypy*.
  Pypy is an alternate python interpreter that performs optimized jitting of python code.

In our tests, it's a 10x speedup out of the box.

- *Only use the SimEngine mixins that you need*. SimEngine uses a mixin model which allows you to add and remove features by constructing new classes. The default engine mixes in every possible features, and the consequence of that is that it is slower than it needs to be. Look at the definition for `UberEngine` (the default SimEngine), copy its declaration, and remove all the base classes which provide features you don't need.
- *Don't load shared libraries unless you need them*.
  The default setting in angr is to try at all costs to find shared libraries that are compatible with the binary you've loaded, including loading them straight out of your OS libraries.
  This can complicate things in a lot of scenarios.
  If you're performing an analysis that's anything more abstract than bare-bones symbolic execution, ESPECIALLY control-flow graph construction, you might want to make the tradeoff of sacrificing accuracy for tractability.
  angr does a reasonable job of making sane things happen when library calls to functions that don't exist try to happen.
- *Use hooking and SimProcedures*.
  If you're enabling shared libraries, then you definitely want to have SimProcedures written for any complicated library function you're jumping into.
  If there's no autonomy requirement for this project, you can often isolate individual problem spots where analysis hangs up and summarize them with a hook.
- *Use SimInspect*.
  SimInspect is the most underused and one of the most powerful features of angr.
  You can hook and modify almost any behavior of angr, including memory index resolution (which is often the slowest part of any angr analysis).
- *Write a concretization strategy*.
  A more powerful solution to the problem of memory index resolution is a concretization strategy.
- *Use the Replacement Solver*.
  You can enable it with the `angr.options.REPLACEMENT_SOLVER` state option.
  The replacement solver allows you to specify AST replacements that are applied at solve-time.
  If you add replacements so that all symbolic data is replaced with concrete data when it comes time to do the solve, the runtime is greatly reduced.
  The API for adding a replacement is `state.se._solver.add_replacement(old, new)`.
  The replacement solver is a bit finicky, so there are some gotchas, but it'll definitely help.

If you're performing lots of concrete or partially-concrete execution

- *Use the unicorn engine*.
  If you have unicorn engine installed, angr can be built to take advantage of it for concrete emulation.
  To enable it, add the options in the set `angr.options.unicorn` to your state.
  Keep in mind that while most items under `angr.options` are individual options, `angr.options.unicorn` is a bundle of options, and is thus a set.
  NOTE: At time of writing the official version of unicorn engine will not work with angr - we have a lot of

patches to it to make it work well with angr.
They're all pending pull requests at this time, so sit tight. If you're really impatient, ping us about uploading our fork!

- *Enable fast memory and fast registers*.
  The state options `angr.options.FAST_MEMORY` and `angr.options.FAST_REGISTERS` will do this.
  These will switch the memory/registers over to a less intensive memory model that sacrifices accuracy for speed.
  TODO: document the specific sacrifices. Should be safe for mostly concrete access though.
  NOTE: not compatible with concretization strategies.
- *Concretize your input ahead of time*.
  This is the approach taken by driller.
  When creating a state with `entry_state` or the like, you can create a SimFile filled with symbolic data, pass it to the initialization function as an argument `entry_state(..., stdin=my_simfile)`, and then constrain the symbolic data in the SimFile to what you want the input to be.
  If you don't require any tracking of the data coming from stdin, you can forego the symbolic part and just fill it with concrete data.
  If there are other sources of input besides standard input, do the same for those.
- *Use the afterburner*. While using unicorn, if you add the `UNICORN_THRESHOLD_CONCRETIZATION` state option, angr will accept thresholds after which it causes symbolic values to be concretized so that execution can spend more time in Unicorn. Specifically, the following thresholds exist:
  - `state.unicorn.concretization_threshold_memory` - this is the number of times a symbolic variable, stored in memory, is allowed to kick execution out of Unicorn before it is forcefully concretized and forced into Unicorn anyways.
  - `state.unicorn.concretization_threshold_registers` - this is the number of times a symbolic variable, stored in a register, is allowed to kick execution out of Unicorn before it is forcefully concretized and forced into Unicorn anyways.
  - `state.unicorn.concretization_threshold_instruction` - this is the number of times that any given instruction can force execution out of Unicorn (by running into symbolic data) before any symbolic data encountered at that instruction is concretized to force execution into Unicorn.

  You can get further control of what is and isn't concretized with the following sets:

  - `state.unicorn.always_concretize` - a set of variable names that will always be concretized to force execution into unicorn (in fact, the memory and register thresholds just end up causing variables to be added to this list).
  - `state.unicorn.never_concretize` - a set of variable names that will never be concretized and forced into Unicorn under any condition.
  - `state.unicorn.concretize_at` - a set of instruction addresses at which data should be concretized and forced into Unicorn. The instruction threshold causes addresses to be added to this set.

  Once something is concretized with the afterburner, you will lose track of that variable. The state will still be consistent, but you'll lose dependencies, as the stuff that comes out of Unicorn is just concrete bits with no memory of what variables they came from. Still, this might be worth it for the speed in some cases, if you know what you want to (or do not want to) concretize.

Memory optimization

The golden rule for memory optimization is to make sure you're not keeping any references to data you don't care about anymore, especially related to states which have been left behind. If you find yourself running out of memory during analysis, the first thing you want to do is make sure you haven't caused a state explosion, meaning that the analysis is accumulating program states too quickly. If the state count is in control, then you can start looking for reference leaks. A good tool to do this with is https://github.com/rhelmot/dumpsterdiver, which gives you an interactive prompt for exploring the reference graph of a python process.

One specific consideration that should be made when analyzing programs with very long paths is that the state history is designed to accumulate data infinitely. This is less of a problem than it could be because the data is stored in a smart tree structure and never copied, but it will accumulate infinitely. To downsize a state's history and free all data related to old steps, call `state.history.trim()`.

One *particularly* problematic member of the history dataset is the basic block trace and the stack pointer trace. When using unicorn engine, these lists of ints can become huge very very quickly. To disable unicorn's capture of ip and sp data, remove the state options `UNICORN_TRACK_BBL_ADDRS` and `UNICORN_TRACK_STACK_POINTERS`.

The Emulated Filesystem

It's very important to be able to control the environment that emulated programs see, including how symbolic data is introduced from the environment! angr has a robust series of abstractions to help you set up the environment you want.

The root of any interaction with the filesystem, sockets, pipes, or terminals is a SimFile object. A SimFile is a *storage* abstraction that defines a sequence of bytes, symbolic or otherwise. There are several kinds of SimFiles which store their data very differently - the two easiest examples are `SimFile` (the base class is actually called `SimFileBase`), which stores files as a flat address-space of data, and `SimPackets`, which stores a sequence of variable-sized reads. The former is best for modeling programs that need to perform seeks on their files, and is the default storage for opened files, while the latter is best for modeling programs that depend on short-reads or use scanf, and is the default storage for stdin/stdout/stderr.

Because SimFiles can have such diverse storage mechanisms, the interface for interacting with them is *very* abstracted. You can read from the file from some position, you can write to the file at some position, you can ask how many bytes are currently stored in the file, and you can concretize the file, generating a testcase for it. If you know specifically which SimFile class you're working with, you can take much more powerful control over it, and as a result you're encouraged to manually create any files you want to work with when you create your initial state.

Specifically, each SimFile class creates its own abstraction of a "position" within the file - each read and write takes a position and returns a new position that you should use to continue from where you left off. If you're working with SimFiles of unknown type you have to treat this position as a totally opaque object with no semantics other than the contract with the read/write functions.

However! This is a very poor match to how programs generally interact with files, so angr also has a SimFileDescriptor abstraction, which provides the familiar read/write/seek/tell interfaces but will also return error conditions when the underlying storage don't support the appropriate operations - just like normal file descriptors!

You may access the mapping from file descriptor number to file descriptor object in `state.posix.fd`. The file descriptor API may be found here.

Just tell me how to do what I want to do!

Okay okay!!

To create a SimFile, you should just create an instance of the class you want to use. Refer to the API docs for the full instructions.

Let's go through a few illustrative examples, which cover how you can work with a concrete file, a symbolic file, a file with mixed concrete and symbolic content, or streams.

Example 1: Create a file with concrete content

```
>>> import angr
>>> simfile = angr.SimFile('myconcretefile', content='hello world!\n')
```

Here's a nuance - you can't use SimFiles without a state attached, because reasons. You'll **never** have to do this in a real scenario (this operation happens automatically when you pass a SimFile into a constructor or the filesystem) but let's mock it up:

```
>>> proj = angr.Project('/bin/true')
>>> state = proj.factory.blank_state()
>>> simfile.set_state(state)
```

To demonstrate the behavior of these files we're going to use the fact that the default SimFile position is just the number of bytes from the start of the file. `SimFile.read` returns a tuple (bitvector data, actual size, new pos):

```
>>> data, actual_size, new_pos = simfile.read(0, 5)
>>> import claripy
>>> assert claripy.is_true(data == 'hello')
>>> assert claripy.is_true(actual_size == 5)
>>> assert claripy.is_true(new_pos == 5)
```

Continue the read, trying to read way too much:

```
>>> data, actual_size, new_pos = simfile.read(new_pos, 1000)
```

angr doesn't try to sanitize the data returned, only the size - we returned 1000 bytes! The intent is that you're only allowed to use up to actual_size of them.

```
>>> assert len(data) == 1000*8  # bitvector sizes are in bits
>>> assert claripy.is_true(actual_size == 8)
>>> assert claripy.is_true(data.get_bytes(0, 8) == ' world!\n')
>>> assert claripy.is_true(new_pos == 13)
```

Example 2: Create a file with symbolic content and a defined size

```
>>> simfile = angr.SimFile('mysymbolicfile', size=0x20)
>>> simfile.set_state(state)

>>> data, actual_size, new_pos = simfile.read(0, 0x30)
>>> assert data.symbolic
>>> assert claripy.is_true(actual_size == 0x20)
```

The basic SimFile provides the same interface as `state.memory`, so you can load data directly:

```
>>> assert simfile.load(0, actual_size) is data.get_bytes(0, 0x20)
```

Example 3: Create a file with constrained symbolic content

```
>>> bytes_list = [claripy.BVS('byte_%d' % i, 8) for i in range(32)]
>>> bytes_ast = claripy.Concat(*bytes_list)
>>> mystate = proj.factory.entry_state(stdin=angr.SimFile('/dev/stdin', content=bytes_ast)
>>> for byte in bytes_list:
...     mystate.solver.add(byte >= 0x20)
...     mystate.solver.add(byte <= 0x7e)
```

Example 4: Create a file with some mixed concrete and symbolic content, but no EOF

```
>>> variable = claripy.BVS('myvar', 10*8)
>>> simfile = angr.SimFile('mymixedfile', content=variable.concat(claripy.BVV('\n')), has_
>>> simfile.set_state(state)
```

We can always query the number of bytes stored in the file:

```
>>> assert claripy.is_true(simfile.size == 11)
```

Reads will generate additional symbolic data past the current frontier:

```
>>> data, actual_size, new_pos = simfile.read(0, 15)
>>> assert claripy.is_true(actual_size == 15)
>>> assert claripy.is_true(new_pos == 15)
>>> assert claripy.is_true(data.get_bytes(0, 10) == variable)
>>> assert claripy.is_true(data.get_bytes(10, 1) == '\n')
>>> assert data.get_bytes(11, 4).symbolic
```

Example 5: Create a file with a symbolic size (`has_end` is implicitly true here)

```
>>> symsize = claripy.BVS('mysize', 64)
>>> state.solver.add(symsize >= 10)
>>> state.solver.add(symsize < 20)
>>> simfile = angr.SimFile('mysymsizefile', size=symsize)
>>> simfile.set_state(state)
```

Reads will encode all possibilities:

```
>>> data, actual_size, new_pos = simfile.read(0, 30)
>>> assert set(state.solver.eval_upto(actual_size, 30)) == set(range(10, 20))
```

The maximum size can't be easily resolved, so the data returned is 30 bytes long, and we're supposed to use it conjunction with actual_size.

```
>>> assert len(data) == 30*8
```

Symbolic read sizes work too!

```
>>> symreadsize = claripy.BVS('myreadsize', 64)
>>> state.solver.add(symreadsize >= 5)
>>> state.solver.add(symreadsize < 30)
>>> data, actual_size, new_pos = simfile.read(0, symreadsize)
```

All sizes between 5 and 20 should be possible:

```
>>> assert set(state.solver.eval_upto(actual_size, 30)) == set(range(5, 20))
```

Example 6: Working with streams (`SimPackets`)

So far, we've only used the SimFile class, which models a random-accessible file object. However, in real life, files are not everything. Streams (standard I/O, TCP, etc.) are a great example: While they hold data like a normal file does, they do not support random accesses, e.g., you cannot read out the second byte of stdin if you have already read passed that position, and you cannot modify any byte that has been previously sent out to a network endpoint. This allows us to design a simpler abstraction for streams in angr.

Believe it or not, this simpler abstraction for streams will benefit symbolic execution. Consider an example program that calls `scanf` N times to read in N strings. With a traditional SimFile, as we do not know the length of each input string, there does not exist any clear boundary in the file between these symbolic input strings. In this case, angr will perform N symbolic reads where each read will generate a gigantic tree of claripy ASTs, with string lengths being symbolic. This is a nightmare for constraint solving. Nevertheless, the fact that `scanf` is used on a stream (stdin) dictates that there will be zero overlap between individual reads, regardless of the sizes of each symbolic input string. We may as well model stdin as a stream that comprises of *consecutive packets*, instead of a file containing a sequence of bytes. Each of the packet can be of a fixed length or a symbolic length. Since there will be absolutely no byte overlap between packets, the constraints that angr will produce after executing this example program will be a lot simpler.

The key concept involved is "short reads", i.e. when you ask for `n` bytes but actually get back fewer bytes than that. We use a different class implementing SimFileBase, `SimPackets`, to automatically enable support for short reads. By default, stdin, stdout, and stderr are all SimPackets objects.

```
>>> simfile = angr.SimPackets('mypackets')
>>> simfile.set_state(state)
```

This'll just generate a single packet. For SimPackets, the position is just a packet number! If left unspecified, short_reads is determined from a state option.

```
>>> data, actual_size, new_pos = simfile.read(0, 20, short_reads=True)
>>> assert len(data) == 20*8
>>> assert set(state.solver.eval_upto(actual_size, 30)) == set(range(21))
```

Data in a SimPackets is stored as tuples of (packet data, packet size) in `.content`.

```
>>> print(simfile.content)
[(<BV160 packet_0_mypackets>, <BV64 packetsize_0_mypackets>)]

>>> simfile.read(0, 1, short_reads=False)
>>> print(simfile.content)
[(<BV160 packet_0_mypackets>, <BV64 packetsize_0_mypackets>), (<BV8 packet_1_mypackets>, <
```

So hopefully you understand sort of the kind of data that a SimFile can store and what'll happen when a program tries to interact with it with various combinations of symbolic and concrete data. Those examples only covered reads, but writes are pretty similar.

The filesystem, for real now

If you want to make a SimFile available to the program, we need to either stick it in the filesystem or serve stdin/stdout from it.

The simulated filesystem is the `state.fs` plugin. You can store, load, and delete files from the filesystem, with the `insert`, `get`, and `delete` methods. Refer to the api docs for details.

So to make our file available as `/tmp/myfile`:

```
>>> state.fs.insert('/tmp/myfile', simfile)
>>> assert state.fs.get('/tmp/myfile') is simfile
```

Then, after execution, we would extract the file from the result state and use `simfile.concretize()` to generate a testcase to reach that state. Keep in mind that `concretize()` returns different types depending on the file type - for a SimFile it's a bytestring and for SimPackets it's a list of bytestrings.

The simulated filesystem supports a fun concept of "mounts", where you can designate a subtree as instrumented by a particular provider. The most common mount is to expose a part of the host filesystem to the guest, lazily importing file data when the program asks for it:

```
>>> state.fs.mount('/', angr.SimHostFilesystem('./guest_chroot'))
```

You can write whatever kind of mount you want to instrument filesystem access by subclassing `angr.SimMount`!

Stdio streams

For stdin and friends, it's a little more complicated. The relevant plugin is `state.posix`, which stores all abstractions relevant to a POSIX-compliant environment. You can always get a state's stdin SimFile with `state.posix.stdin`, but you can't just replace it - as soon as the state is created, references to this file are created in the file descriptors. Because of this you need to specify it at the time the POSIX plugin is created:

```
>>> state.register_plugin('posix', angr.state_plugins.posix.SimSystemPosix(stdin=simfile, 
>>> assert state.posix.stdin is simfile
>>> assert state.posix.stdout is simfile
>>> assert state.posix.stderr is simfile
```

Or, there's a nice shortcut while creating the state if you only need to specify stdin:

```
>>> state = proj.factory.entry_state(stdin=simfile)
>>> assert state.posix.stdin is simfile
```

Any of those places you can specify a SimFileBase, you can also specify a string or a bitvector (a flat SimFile with fixed size will be created to hold it) or a SimFile type (it'll be instantiated for you).

Intermediate Representation

In order to be able to analyze and execute machine code from different CPU architectures, such as MIPS, ARM, and PowerPC in addition to the classic x86, angr performs most of its analysis on an *intermediate representation*, a structured description of the fundamental actions performed by each CPU instruction. By understanding angr's IR, VEX (which we borrowed from Valgrind), you will be able to write very quick static analyses and have a better understanding of how angr works.

The VEX IR abstracts away several architecture differences when dealing with different architectures, allowing a single analysis to be run on all of them:

- **Register names.** The quantity and names of registers differ between architectures, but modern CPU designs hold to a common theme: each CPU contains several general purpose registers, a register to hold the stack pointer, a set of registers to store condition flags, and so forth. The IR provides a consistent, abstracted interface to registers on different platforms. Specifically, VEX models the registers as a separate memory space, with integer offsets (e.g., AMD64's `rax` is stored starting at address 16 in this memory space).

- **Memory access.** Different architectures access memory in different ways. For example, ARM can access memory in both little-endian and big-endian modes. The IR abstracts away these differences.
- **Memory segmentation.** Some architectures, such as x86, support memory segmentation through the use of special segment registers. The IR understands such memory access mechanisms.
- **Instruction side-effects.** Most instructions have side-effects. For example, most operations in Thumb mode on ARM update the condition flags, and stack push/pop instructions update the stack pointer. Tracking these side-effects in an *ad hoc* manner in the analysis would be crazy, so the IR makes these effects explicit.

There are lots of choices for an IR. We use VEX, since the uplifting of binary code into VEX is quite well supported. VEX is an architecture-agnostic, side-effects-free representation of a number of target machine languages. It abstracts machine code into a representation designed to make program analysis easier. This representation has four main classes of objects:

- **Expressions.** IR Expressions represent a calculated or constant value. This includes memory loads, register reads, and results of arithmetic operations.
- **Operations.** IR Operations describe a *modification* of IR Expressions. This includes integer arithmetic, floating-point arithmetic, bit operations, and so forth. An IR Operation applied to IR Expressions yields an IR Expression as a result.
- **Temporary variables.** VEX uses temporary variables as internal registers: IR Expressions are stored in temporary variables between use. The content of a temporary variable can be retrieved using an IR Expression. These temporaries are numbered, starting at `t0`. These temporaries are strongly typed (e.g., "64-bit integer" or "32-bit float").
- **Statements.** IR Statements model changes in the state of the target machine, such as the effect of memory stores and register writes. IR Statements use IR Expressions for values they may need. For example, a memory store *IR Statement* uses an *IR Expression* for the target address of the write, and another *IR Expression* for the content.
- **Blocks.** An IR Block is a collection of IR Statements, representing an extended basic block (termed "IR Super Block" or "IRSB") in the target architecture. A block can have several exits. For conditional exits from the middle of a basic block, a special *Exit* IR Statement is used. An IR Expression is used to represent the target of the unconditional exit at the end of the block.

VEX IR is actually quite well documented in the `libvex_ir.h` file (https://github.com/angr/vex/blob/master/pub/libvex_ir.h) in the VEX repository. For the lazy, we'll detail some parts of VEX that you'll likely interact with fairly frequently. To begin with, here are some IR Expressions:

IR Expression	Evaluated Value	VEX Output Example

Constant	A constant value.	0x4:I32
Read Temp	The value stored in a VEX temporary variable.	RdTmp(t10)
Get Register	The value stored in a register.	GET:I32(16)
Load Memory	The value stored at a memory address, with the address specified by another IR Expression.	LDle:I32 / LDbe:I64
Operation	A result of a specified IR Operation, applied to specified IR Expression arguments.	Add32
If-Then-Else	If a given IR Expression evaluates to 0, return one IR Expression. Otherwise, return another.	ITE
Helper Function	VEX uses C helper functions for certain operations, such as computing the conditional flags registers of certain architectures. These functions return IR Expressions.	function_name()

These expressions are then, in turn, used in IR Statements. Here are some common ones:

IR Statement	Meaning	VEX Output Example
Write Temp	Set a VEX temporary variable to the value of the given IR Expression.	WrTmp(t1) = (IR Expression)
Put Register	Update a register with the value of the given IR Expression.	PUT(16) = (IR Expression)
Store Memory	Update a location in memory, given as an IR Expression, with a value, also given as an IR Expression.	STle(0x1000) = (IR Expressior
Exit	A conditional exit from a basic block, with the jump target specified by an IR Expression. The condition is specified by an IR Expression.	if (condition) goto (Boring) 0x4000A00:I32

An example of an IR translation, on ARM, is produced below. In the example, the subtraction operation is translated into a single IR block comprising 5 IR Statements, each of which contains at least one IR Expression (although, in real life, an IR block would typically consist of more than one instruction). Register names are translated into numerical indices given to the *GET* Expression and *PUT* Statement. The astute reader will observe that the actual subtraction is modeled by the first 4 IR Statements of the block, and the incrementing of the program counter to point to the next instruction (which, in this case, is located at `0x59FC8`) is modeled by the last statement.

The following ARM instruction:

```
subs R2, R2, #8
```

Becomes this VEX IR:

```
t0 = GET:I32(16)
t1 = 0x8:I32
t3 = Sub32(t0,t1)
PUT(16) = t3
PUT(68) = 0x59FC8:I32
```

Now that you understand VEX, you can actually play with some VEX in angr: We use a library called PyVEX that exposes VEX into Python. In addition, PyVEX implements its own pretty-printing so that it can show register names instead of register offsets in PUT and GET instructions.

PyVEX is accessable through angr through the `Project.factory.block` interface. There are many different representations you could use to access syntactic properties of a block of code, but they all have in common the trait of analyzing a particular sequence of bytes. Through the `factory.block` constructor, you get a `Block` object that can be easily turned into several different representations. Try `.vex` for a PyVEX IRSB, or `.capstone` for a Capstone block.

Let's play with PyVEX:

```
>>> import angr

# load the program binary
>>> proj = angr.Project("/bin/true")

# translate the starting basic block
>>> irsb = proj.factory.block(proj.entry).vex
# and then pretty-print it
>>> irsb.pp()

# translate and pretty-print a basic block starting at an address
>>> irsb = proj.factory.block(0x401340).vex
>>> irsb.pp()

# this is the IR Expression of the jump target of the unconditional exit at the end of the
```

```
>>> print(irsb.next)
# this is the type of the unconditional exit (e.g., a call, ret, syscall, etc)
>>> print(irsb.jumpkind)

# you can also pretty-print it
>>> irsb.next.pp()

# iterate through each statement and print all the statements
>>> for stmt in irsb.statements:
...     stmt.pp()

# pretty-print the IR expression representing the data, and the *type* of that IR expressi
>>> import pyvex
>>> for stmt in irsb.statements:
...     if isinstance(stmt, pyvex.IRStmt.Store):
...         print("Data:",)
...         stmt.data.pp()
...         print("")
...         print("Type:",)
...         print(stmt.data.result_type)
...         print("")

# pretty-print the condition and jump target of every conditional exit from the basic bloc
>>> for stmt in irsb.statements:
...     if isinstance(stmt, pyvex.IRStmt.Exit):
...         print("Condition:",)
...         stmt.guard.pp()
...         print("")
...         print("Target:",)
...         stmt.dst.pp()
...         print("")

# these are the types of every temp in the IRSB
>>> print(irsb.tyenv.types)

# here is one way to get the type of temp 0
>>> print(irsb.tyenv.types[0])
```

Condition flags computation (for x86 and ARM)

One of the most common instruction side-effects on x86 and ARM CPUs is updating condition flags, such as the zero flag, the carry flag, or the overflow flag. Computer architects usually put the concatenation of these flags (yes, concatenation of the flags, since each condition flag is 1 bit wide) into a special register (i.e. `EFLAGS` / `RFLAGS` on x86, `APSR` / `CPSR` on ARM). This special register stores important information about the program state, and is critical for correct emulation of the CPU.

VEX uses 4 registers as its "Flag thunk descriptors" to record details of the latest flag-setting operation. VEX has a lazy strategy to compute the flags: when an operation that would update the flags happens, instead of computing the flags, VEX stores a code representing this operation to the `cc_op` pseudo-register, and the arguments to the operation in `cc_dep1` and `cc_dep2`. Then, whenever VEX needs to get the actual flag values, it can figure out what the one bit corresponding to the flag in question actually is, based on its flag thunk descriptors. This is an optimization in the flags computation, as VEX can now just directly perform the relevant operation in the IR without bothering to compute and update the flags' value.

Amongst different operations that can be placed in `cc_op`, there is a special value 0 which corresponds to `OP_COPY` operation. This operation is supposed to copy the value in `cc_dep1` to the flags. It simply means that `cc_dep1` contains the flags' value. angr uses this fact to let us efficiently retrieve the flags' value: whenever we ask for the actual flags, angr computes their value, then dumps them back into `cc_dep1` and sets `cc_op = OP_COPY` in order to cache the computation. We can also use this operation to allow the user to write to the flags: we just set `cc_op = OP_COPY` to say that a new value being set to the flags, then set `cc_dep1` to that new value.

Working with Data and Conventions

Frequently, you'll want to access structured data from the program you're analyzing. angr has several features to make this less of a headache.

Working with types

angr has a system for representing types. These SimTypes are found in `angr.types` - an instance of any of these classes represents a type. Many of the types are incomplete unless they are supplamented with a SimState - their size depends on the architecture you're running under. You may do this with `ty.with_arch(arch)`, which returns a copy of itself, with the architecture specified.

angr also has a light wrapper around `pycparser`, which is a C parser. This helps with getting instances of type objects:

```
>>> import angr, monkeyhex

# note that SimType objects have their __repr__ defined to return their c type name,
# so this function actually returned a SimType instance.
>>> angr.types.parse_type('int')
int

>>> angr.types.parse_type('char **')
char**

>>> angr.types.parse_type('struct aa {int x; long y;}')

struct aa

>>> angr.types.parse_type('struct aa {int x; long y;}').fields
OrderedDict([('x', int), ('y', long)])
```

Additionally, you may parse C definitions and have them returned to you in a dict, either of variable/function declarations or of newly defined types:

```
>>> angr.types.parse_defns("int x; typedef struct llist { char* str; struct llist *next; }
{'x': int, 'y': struct llist*}

>>> defs = angr.types.parse_types("int x; typedef struct llist { char* str; struct llist *
>>> defs
{'struct llist': struct llist, 'list_node': struct llist}

# if you want to get both of these dicts at once, use parse_file, which returns both in a
>>> angr.types.parse_file("int x; typedef struct llist { char* str; struct llist *next; }
({'x': int, 'y': struct llist*},
 {'struct llist': struct llist, 'list_node': struct llist})

>>> defs['list_node'].fields
OrderedDict([('str', char*), ('next', struct llist*)])

>>> defs['list_node'].fields['next'].pts_to.fields
OrderedDict([('str', char*), ('next', struct llist*)])

# If you want to get a function type and you don't want to construct it manually,
# you can use parse_type
>>> angr.types.parse_type("int (int y, double z)")
(int, double) -> int
```

And finally, you can register struct definitions for future use:

```
>>> angr.types.register_types(angr.types.parse_type('struct abcd { int x; int y; }'))
>>> angr.types.register_types(angr.types.parse_types('typedef long time_t;'))
>>> angr.types.parse_defns('struct abcd a; time_t b;')
{'a': struct abcd, 'b': long}
```

These type objects aren't all that useful on their own, but they can be passed to other parts of angr to specify data types.

Accessing typed data from memory

Now that you know how angr's type system works, you can unlock the full power of the `state.mem` interface! Any type that's registered with the types module can be used to extract data from memory.

```
>>> p = angr.Project('examples/fauxware/fauxware')
>>> s = p.factory.entry_state()
>>> s.mem[0x601048]
<<untyped> <unresolvable> at 0x601048>

```

```
>>> s.mem[0x601048].long
<long (64 bits) <BV64 0x4008d0> at 0x601048>

>>> s.mem[0x601048].long.resolved
<BV64 0x4008d0>

>>> s.mem[0x601048].long.concrete
0x4008d0

>>> s.mem[0x601048].struct.abcd
<struct abcd {
  .x = <BV32 0x4008d0>,
  .y = <BV32 0x0>
} at 0x601048>

>>> s.mem[0x601048].struct.abcd.x
<int (32 bits) <BV32 0x4008d0> at 0x601048>

>>> s.mem[0x601048].struct.abcd.y
<int (32 bits) <BV32 0x0> at 0x60104c>

>>> s.mem[0x601048].deref
<<untyped> <unresolvable> at 0x4008d0>

>>> s.mem[0x601048].deref.string
<string_t <BV64 0x534f534e45414b59> at 0x4008d0>

>>> s.mem[0x601048].deref.string.resolved
<BV64 0x534f534e45414b59>

>>> s.mem[0x601048].deref.string.concrete
b'SOSNEAKY'
```

The interface works like this:

- You first use [array index notation] to specify the address you'd like to load from
- If at that address is a pointer, you may access the `deref` property to return a SimMemView at the address present in memory.
- You then specify a type for the data by simply accessing a property of that name. For a list of supported types, look at `state.mem.types`.
- You can then *refine* the type. Any type may support any refinement it likes. Right now the only refinements supported are that you may access any member of a struct by its member name, and you may index into a string or array to access that element.
- If the address you specified initially points to an array of that type, you can say `.array(n)` to view the data as an array of n elements.
- Finally, extract the structured data with `.resolved` or `.concrete`. `.resolved` will return bitvector values, while `.concrete` will return integer, string, array, etc values, whatever best represents the data.
- Alternately, you may store a value to memory, by assigning to the chain of properties that you've

constructed. Note that because of the way python works, `x = s.mem[...].prop; x = val` will NOT work, you must say `s.mem[...].prop = val`.

If you define a struct using `register_types(parse_type(struct_expr))`, you can access it here as a type:

```
>>> s.mem[p.entry].struct.abcd
<struct abcd {
  .x = <BV32 0x8949ed31>,
  .y = <BV32 0x89485ed1>
} at 0x400580>
```

Working with Calling Conventions

A calling convention is the specific means by which code passes arguments and return values through function calls. angr's abstraction of calling conventions is called SimCC. You can construct new SimCC instances through the angr object factory, with `p.factory.cc(...)`. This will give a calling convention which is guessed based your guest architecture and OS. If angr guesses wrong, you can explicitly pick one of the calling conventions in the `angr.calling_conventions` module.

If you have a very wacky calling convention, you can use `angr.calling_conventions.SimCCUsercall`. This will ask you to specify locations for the arguments and the return value. To do this, use instances of the `SimRegArg` or `SimStackArg` classes. You can find them in the factory - `p.factory.cc.Sim*Arg`.

Once you have a SimCC object, you can use it along with a SimState object and a function prototype (a SimTypeFunction) to extract or store function arguments more cleanly. Take a look at the API documentation for details. Alternately, you can pass it to an interface that can use it to modify its own behavior, like `p.factory.call_state`, or...

Callables

Callables are a Foreign Functions Interface (FFI) for symbolic execution. Basic callable usage is to create one with `myfunc = p.factory.callable(addr)`, and then call it! `result = myfunc(args, ...)` When you call the callable, angr will set up a `call_state` at the given address, dump the given arguments into memory, and run a `path_group` based on this state until all the paths have exited from the function. Then, it merges all the result states together, pulls the return value out of that state, and returns it.

All the interaction with the state happens with the aid of a `SimCC` and a `SimTypeFunction`, to tell where to put the arguments and where to get the return value. It will try to use a sane default for the architecture, but if you'd like to customize it, you can pass a `SimCC` object in the `cc` keyword argument when constructing the callable. The `SimTypeFunction` is required - you must pass the `prototype` parameter. If you pass a string to this parameter it will be parsed as a function declaration.

You can pass symbolic data as function arguments, and everything will work fine. You can even pass more complicated data, like strings, lists, and structures as native python data (use tuples for structures), and it'll be serialized as cleanly as possible into the state. If you'd like to specify a pointer to a certain value, you can wrap it in a `PointerWrapper` object, available as `p.factory.callable.PointerWrapper`. The exact semantics of how pointer-wrapping work are a little confusing, but they can be boiled down to "unless you specify it with a PointerWrapper or a specific SimArrayType, nothing will be wrapped in a pointer automatically unless it gets to the end and it hasn't yet been wrapped in a pointer yet and the original type is a string, array, or tuple." The relevant code is actually in SimCC - it's the `setup_callsite` function.

If you don't care for the actual return value of the call, you can say `func.perform_call(arg, ...)`, and then the properties `func.result_state` and `func.result_path_group` will be populated. They will actually be populated even if you call the callable normally, but you probably care about them more in this case!

Claripy

angr's solver engine is called Claripy. Claripy exposes the following design:

- Claripy ASTs (the subclasses of claripy.ast.Base) provide a unified way to interact with concrete and symbolic expressions
- `Frontend`s provide different paradigms for evaluating these expressions. For example, the `FullFrontend` solves expressions using something like an SMT solver backend, while `LightFrontend` handles them by using an abstract (and approximating) data domain backend.
- Each `Frontend` needs to, at some point, do actual operation and evaluations on an AST. ASTs don't support this on their own. Instead, `Backend`s translate ASTs into backend objects (i.e., python primitives for `BackendConcrete`, Z3 expressions for `BackendZ3`, strided intervals for `BackendVSA`, etc) and handle any appropriate state-tracking objects (such as tracking the solver state in the case of `BackendZ3`). Roughly speaking, frontends take ASTs as inputs and use backends to `backend.convert()` those ASTs into backend objects that can be evaluated and otherwise reasoned about.
- `FrontendMixin`s customize the operation of `Frontend`s. For example, `ModelCacheMixin` caches solutions from an SMT solver.
- The combination of a Frontend, a number of FrontendMixins, and a number of Backends comprise a claripy `Solver`.

Internally, Claripy seamlessly mediates the co-operation of multiple disparate backends -- concrete bitvectors, VSA constructs, and SAT solvers. It is pretty badass.

Most users of angr will not need to interact directly with Claripy (except for, maybe, claripy AST objects, which represent symbolic expressions) -- angr handles most interactions with Claripy internally. However, for dealing with expressions, an understanding of Claripy might be useful.

Claripy ASTs

Claripy ASTs abstract away the differences between mathematical constructs that Claripy supports. They define a tree of operations (i.e., `(a + b) / c)` on any type of underlying data. Claripy handles the application of these operations on the underlying objects themselves by dispatching requests to the backends.

Currently, Claripy supports the following types of ASTs:

Name	Description	Supported By (Claripy Backends)	Example Code
BV	This is a bitvector, whether symbolic (with a name) or concrete (with a value). It has a size (in bits).	BackendConcrete, BackendVSA, BackendZ3	Create a 32-bit symbolic bitvector "x" `claripy.BVS('x', 32)` Create a 32-bit bitvector with the valu `0xc001b3475`: `claripy.BVV(0xc001b a75, 32)` Create a 32-bit "stride interval" (see VSA documentation) that can be any divisible-by-10 number betwee 1000 and 2000: `claripy.SI(name='x', bits=32, lower_bound=1000, upper_bound=2000, stride=10)`
FP	This is a floating-point number, whether symbolic (with a name) or concrete (with a value).	BackendConcrete, BackendZ3	Create a `claripy.fp.FSORT_D UBLE` symbolic floating point "b": `claripy.FPS('b', claripy.fp.FSORT_D UBLE)` Create a `claripy.fp.FSORT_FL OAT` floating point wi value `3.2`: `claripy.FPV(3.2, claripy.fp.FSORT_FL AT)`
			`claripy.BoolV(True)`, or

Bool	This is a boolean operation (True or False).	BackendConcrete, BackendVSA, BackendZ3	`claripy.true` or `claripy.false`, c by comparing two ASTs (i.e., `claripy.BVS('x'` `32) <` `claripy.BVS('y'` `32)`

All of the above creation code returns claripy.AST objects, on which operations can then be carried out.

ASTs provide several useful operations.

```
>>> import claripy

>>> bv = claripy.BVV(0x41424344, 32)

# Size - you can get the size of an AST with .size()
>>> assert bv.size() == 32

# Reversing - .reversed is the reversed version of the BVV
>>> assert bv.reversed is claripy.BVV(0x44434241, 32)
>>> assert bv.reversed.reversed is bv

# Depth - you can get the depth of the AST
>>> print(bv.depth)
>>> assert bv.depth == 1
>>> x = claripy.BVS('x', 32)
>>> assert (x+bv).depth == 2
>>> assert ((x+bv)/10).depth == 3
```

Applying a condition (==, !=, etc) on ASTs will return an AST that represents the condition being carried out. For example:

```
>>> r = bv == x
>>> assert isinstance(r, claripy.ast.Bool)


>>> p = bv == bv
>>> assert isinstance(p, claripy.ast.Bool)
>>> assert p.is_true()
```

You can combine these conditions in different ways.

```
>>> q = claripy.And(claripy.Or(bv == x, bv * 2 == x, bv * 3 == x), x == 0)
>>> assert isinstance(p, claripy.ast.Bool)
```

The usefulness of this will become apparent when we discuss Claripy solvers.

In general, Claripy supports all of the normal python operations (+, -, |, ==, etc), and provides additional ones via the Claripy instance object. Here's a list of available operations from the latter.

Name	Description	Example
LShR	Logically shifts a bit expression (BVV, BV, SI) to the right.	`claripy.LShR(x, 10)`
SignExt	Sign-extends a bit expression.	`claripy.SignExt(32, x` or `x.sign_extend(32)`
ZeroExt	Zero-extends a bit expression.	`claripy.ZeroExt(32, x` or `x.zero_extend(32)`
Extract	Extracts the given bits (zero-indexed from the *right*, inclusive) from a bit expression.	Extract the rightmost byte of x: `claripy.Extract(7, 0, x)` or `x[7:0]`
Concat	Concatenates several bit expressions together into a new bit expression.	`claripy.Concat(x, y, z)`
RotateLeft	Rotates a bit expression left.	`claripy.RotateLeft(x, 8)`
RotateRight	Rotates a bit expression right.	`claripy.RotateRight(x 8)`
Reverse	Endian-reverses a bit expression.	`claripy.Reverse(x)` or `x.reversed`
And	Logical And (on boolean expressions)	`claripy.And(x == y, x > 0)`
Or	Logical Or (on boolean expressions)	`claripy.Or(x == y, y 10)`
Not	Logical Not (on a boolean expression)	`claripy.Not(x == y)` is the same as `x != y`
If	An If-then-else	Choose the maximum of two expressions: `claripy.If(x > y, x, y)`
ULE	Unsigned less than or equal to.	Check if x is less than or equal to y: `claripy.ULE(x, y)`
ULT	Unsigned less than.	Check if x is less than y: `claripy.ULT(x, y)`
UGE	Unsigned greater than or equal	Check if x is greater than or equal to y: `claripy.UGE(x,`

	to.	y)
UGT	Unsigned greater than.	Check if x is greater than y: `claripy.UGT(x, y)`
SLE	Signed less than or equal to.	Check if x is less than or equal to y: `claripy.SLE(x, y)`
SLT	Signed less than.	Check if x is less than y: `claripy.SLT(x, y)`
SGE	Signed greater than or equal to.	Check if x is greater than or equal to y: `claripy.SGE(x, y)`

NOTE: The default python `>`, `<`, `>=`, and `<=` are unsigned in Claripy. This is different than their behavior in Z3, because it seems more natural in binary analysis.

Solvers

The main point of interaction with Claripy are the Claripy Solvers. Solvers expose an API to interpret ASTs in different ways and return usable values. There are several different solvers.

Name	Description
Solver	This is analogous to a `z3.Solver()`. It is a solver that tracks constraints on symbolic variable and uses a constraint solver (currently, Z3) to evaluate symbolic expressions.
SolverVSA	This solver uses VSA to reason about values. It is an *approximating* solver, but produces values without performing actual constraint solves.
SolverReplacement	This solver acts as a pass-through to a child solve allowing the replacement of expressions on-the-fl It is used as a helper by other solvers and can be used directly to implement exotic analyses.
SolverHybrid	This solver combines the SolverReplacement and the Solver (VSA and Z3) to allow for *approximatin* values. You can specify whether or not you want a exact result from your evaluations, and this solver does the rest.
	This solver implements optimizations that solve

SolverComposite	smaller sets of constraints to speed up constraint solving.

Some examples of solver usage:

```
# create the solver and an expression
>>> s = claripy.Solver()
>>> x = claripy.BVS('x', 8)

# now let's add a constraint on x
>>> s.add(claripy.ULT(x, 5))

>>> assert sorted(s.eval(x, 10)) == [0, 1, 2, 3, 4]
>>> assert s.max(x) == 4
>>> assert s.min(x) == 0

# we can also get the values of complex expressions
>>> y = claripy.BVV(65, 8)
>>> z = claripy.If(x == 1, x, y)
>>> assert sorted(s.eval(z, 10)) == [1, 65]

# and, of course, we can add constraints on complex expressions
>>> s.add(z % 5 != 0)
>>> assert s.eval(z, 10) == (1,)
>>> assert s.eval(x, 10) == (1,) # interestingly enough, since z can't be y, x can only be
```

Custom solvers can be built by combining a Claripy Frontend (the class that handles the actual interaction with SMT solver or the underlying data domain) and some combination of frontend mixins (that handle things like caching, filtering out duplicate constraints, doing opportunistic simplification, and so on).

Claripy Backends

Backends are Claripy's workhorses. Claripy exposes ASTs to the world, but when actual computation has to be done, it pushes those ASTs into objects that can be handled by the backends themselves. This provides a unified interface to the outside world while allowing Claripy to support different types of computation. For example, BackendConcrete provides computation support for concrete bitvectors and booleans, BackendVSA introduces VSA constructs such as StridedIntervals (and details what happens when operations are performed on them, and BackendZ3 provides support for symbolic variables and constraint solving.

There are a set of functions that a backend is expected to implement. For all of these functions, the "public" version is expected to be able to deal with claripy's AST objects, while the "private" version should only deal with objects specific to the backend itself. This is distinguished with Python idioms: a public function will be named func() while a private function will be _func(). All functions should return objects that are usable by the backend in its private methods. If this can't be done (i.e., some functionality is being attempted that the backend can't handle), the backend should raise a BackendError. In this case, Claripy will move on to the next backend in its list.

All backends must implement a `convert()` function. This function receives a claripy AST and should return an object that the backend can handle in its private methods. Backends should also implement a `_convert()` method, which will receive anything that is *not* a claripy AST object (i.e., an integer or an object from a different backend). If `convert()` or `_convert()` receives something that the backend can't translate to a format that is usable internally, the backend should raise BackendError, and thus won't be used for that object. All backends must also implement any functions of the base `Backend` abstract class that currently raise `NotImplementedError()`.

Claripy's contract with its backends is as follows: backends should be able to handle, in their private functions, any object that they return from their private *or* public functions. Claripy will never pass an object to any backend private function that did not originate as a return value from a private or public function of that backend. One exception to this is `convert()` and `_convert()`, as Claripy can try to stuff anything it feels like into _convert() to see if the backend can handle that type of object.

Backend Objects

To perform actual, useful computation on ASTs, Claripy uses backend objects. A `BackendObject` is a result of the operation represented by the AST. Claripy expects these objects to be returned from their respective backends, and will pass such objects into that backend's other functions.

Symbolic Memory Addressing

angr supports *symbolic memory addressing*, meaning that offsets into memory may be symbolic. Our implementation of this is inspired by "Mayhem". Specifically, this means that angr concretizes symbolic addresses when they are used as the target of a write. This causes some surprises, as users tend to expect symbolic writes to be treated purely symbolically, or "as symbolically" as we treat symbolic reads, but that is not the default behavior. However, like most things in angr, this is configurable.

The address resolution behavior is governed by *concretization strategies*, which are subclasses of `angr.concretization_strategies.SimConcretizationStrategy`. Concretization strategies for reads are set in `state.memory.read_strategies` and for writes in `state.memory.write_strategies`. These strategies are called, in order, until one of them is able to resolve addresses for the symbolic index. By setting your own concretization strategies (or through the use of SimInspect `address_concretization` breakpoints, described above), you can change the way angr resolves symbolic addresses.

For example, angr's default concretization strategies for writes are:

1. A conditional concretization strategy that allows symbolic writes (with a maximum range of 128 possible solutions) for any indices that are annotated with `angr.plugins.symbolic_memory.MultiwriteAnnotation`.
2. A concretization strategy that simply selects the maximum possible solution of the symbolic index.

To enable symbolic writes for all indices, you can either add the `SYMBOLIC_WRITE_ADDRESSES` state option at state creation time or manually insert a `angr.concretization_strategies.SimConcretizationStrategyRange` object into `state.memory.write_strategies`. The strategy object takes a single argument, which is the maximum range of possible solutions that it allows before giving up and moving on to the next (presumably non-symbolic) strategy.

Writing concretization strategies

TODO

Java Symbolic Execution

`angr` also supports symbolically executing Java code and Android apps! This also includes Android apps using a combination of compiled Java and native (C/C++) code.

Java support is experimental! *Contribution from the community is highly encouraged! Pull requests are very welcomed!*

We implemented Java support by lifting the compiled Java code, both Java and DEX bytecode, leveraging our Soot python wrapper: pysoot. `pysoot` extracts a fully serializable interface from Android apps and Java code (unfortunately, as of now, it only works on Linux). For every class of the generated IR (for instance, `SootMethod`), you can nicely print its instructions (in a format similar to `Soot` `shimple`) using `print()` or `str()`.

We then leverage the generated IR in a new angr engine able to run code in Soot IR: angr/engines/soot/engine.py. This engine is also able to automatically switch to executing native code if the Java code calls any native method using the JNI interface.

Together with the symbolic execution, we also implemented some basic static analysis, specifically a basic CFG reconstruction analysis. Moreover, we added support for string constraint solving, modifying claripy and using the CVC4 solver.

How to install

Enabling Java support requires few more steps than typical angr installation. Assuming you installed angr-dev, activate the virtualenv and run:

```
# CVC4 and pysoot should be already installed (if you used angr-dev to install angr)
# install cvc4, needed for String solving
pip install cvc4-solver
# install pysoot, needed to lift code from JARs and APKs
git clone git@github.com:angr/pysoot.git
cd pysoot
pip install -e .
cd ..
```

```
# install a specific version of pysmt (the one currently available on pip is buggy)
pip uninstall pysmt
git clone https://github.com/pysmt/pysmt.git
cd pysmt
git checkout 6d792db47be5f8734db15848faca9bc6b770085e
pip install -e .
cd ..
```

Analyzing Android apps.

Analyzing Android apps (`.APK` files, containing Java code compiled to the `DEX` format) requires the Android SDK. Typically, it is installed in `<HOME>/Android/SDK/platforms/platform-XX/android.jar`, where `XX` is the Android SDK version used by the app you want to analyze (you may want to install all the platforms required by the Android apps you want to analyze).

Examples

There are multiple examples available:

- Easy Java crackmes: java_crackme1, java_simple3, java_simple4
- A more complex example (solving a CTF challenge): ictf2017_javaisnotfun, blogpost
- Symbolically executing an Android app (using a mix of Java and native code): java_androidnative1
- Many other low-level tests: test_java

Symbion

Let's suppose you want to symbolically analyze a specific function of a program, but there is a huge initialization step that you want to skip because it is not necessary for your analysis, or cannot properly be emulated by angr. For example, maybe your program is running on an embedded system and you have access to a debug interface, but you can't easily replicate the hardware in a simulated environment.

This is the perfect scenario for `Symbion`, our interleaved execution technique!

We implemented a built-in system that let users define a `ConcreteTarget` that is used to "import" a concrete state of the target program from an external source into `angr`. Once the state is imported you can make parts of the state symbolic, use symbolic execution on this state, run your analyses, and finally concretize the symbolic parts and resume concrete execution in the external environment. By iterating this process it is possible to implement run-time and interactive advanced symbolic analyses that are backed up by the real program's execution!

Isn't that cool?

How to install

To use this technique you'll need an implementation of a `ConcreteTarget` (effectively, an object that is going to be the "glue" between angr and the external process.) We ship a default one (the AvatarGDBConcreteTarget, which control an instance of a program being debugged under GDB) in the following repo https://github.com/angr/angr-targets.

Assuming you installed angr-dev, activate the virtualenv and run:

```
git clone https://github.com/angr/angr-targets.git
cd angr-targets
pip install .
```

Now you're ready to go!

Gists

Once you have created an entry state, instantiated a `SimulationManager`, and specified a list of *stop_points* using the `Symbion` interface we are going to resume the concrete process execution.

```python
# Instantiating the ConcreteTarget
avatar_gdb = AvatarGDBConcreteTarget(avatar2.archs.x86.X86_64,
                                     GDB_SERVER_IP, GDB_SERVER_PORT)

# Creating the Project
p = angr.Project(binary_x64, concrete_target=avatar_gdb,
                             use_sim_procedures=True)

# Getting an entry_state
entry_state = p.factory.entry_state()

# Forget about these options as for now, will explain later.
entry_state.options.add(angr.options.SYMBION_SYNC_CLE)
entry_state.options.add(angr.options.SYMBION_KEEP_STUBS_ON_SYNC)

# Use Symbion!
simgr.use_technique(angr.exploration_techniques.Symbion(find=[0x85b853])
```

When one of your stop_points (effectively a breakpoint) is hit, we give control to `angr`. A new plugin called *concrete* is in charge of synchronizing the concrete state of the program inside a new `SimState`.

Roughly, synchronization does the following:

- All the registers' values (NOT marked with concrete=False in the respective arch file in archinfo) are copied inside the new SimState.
- The underlying memory backend is hooked in a way that all the further memory accesses triggered

during symbolic execution are redirected to the concrete process.

- If the project is initialized with SimProcedure (use_sim_procedures=True) we are going to re-hook the external functions' addresses with a `SimProcedure` if we happen to have it, otherwise with a `SimProcedure` stub (you can control this decision by using the Options SYMBION_KEEP_STUBS_ON_SYNC). Conversely, the real code of the function is executed inside angr (Warning: do that at your own risk!)

Once this process is completed, you can play with your new `SimState` backed by the concrete process stopped at that particular stop_point. Options

The way we synchronize the concrete process inside angr is customizable by 2 state options:

- **SYMBION_SYNC_CLE**: this option controls the synchronization of the memory mapping of the program inside angr. When the project is created, the memory mapping inside angr is different from the one inside the concrete process (this will change as soon as Symbion will be fully compatible with archr). If you want the process mapping to be fully synchronized with the one of the concrete process, set this option to the SimState before initializing the SimulationManager (Note that this is going to happen at the first synchronization of the concrete process inside angr, NOT before)

```
entry_state.options.add(angr.options.SYMBION_SYNC_CLE)
simgr = project.factory.simgr(state)
```

- **SYMBION_KEEP_STUBS_ON_SYNC**: this option controls how we re-hook external functions with SimProcedures. If the project has been initialized to use SimProcedures (use_sim_procedures=True), we are going to re-hook external functions with SimProcedures (if we have that particular implementation) or with a generic stub. If you want to execute SimProcedures for functions for which we have an available implementation and a generic stub SimProcedure for the ones we have not, set this option to the SimState before initializing the SimulationManager. In the other case, we are going to execute the real code for the external functions that miss a SimProcedure (no generic stub is going to be used).

```
entry_state.options.add(angr.options.SYMBION_KEEP_STUBS_ON_SYNC)
simgr = project.factory.simgr(state)
```

Example
You can find more information about this technique and a complete example in our blog post: https://angr.io/blog/angr_symbion/. For more technical details a public paper will be available soon, or, ping @degrigis on our `angr` Slack channel.

Extending angr

Programming SimProcedures

Hooks in angr are very powerful! You can use them to modify a program's behavior in any way you could imagine. However, the exact way you might want to program a specific hook may be non-obvious. This chapter should serve as a guide when programming SimProcedures.

Quick Start

Here's an example that will remove all bugs from any program:

```
>>> from angr import Project, SimProcedure
>>> project = Project('examples/fauxware/fauxware')

>>> class BugFree(SimProcedure):
...    def run(self, argc, argv):
...        print('Program running with argc=%s and argv=%s' % (argc, argv))
...        return 0

# this assumes we have symbols for the binary
>>> project.hook_symbol('main', BugFree())

# Run a quick execution!
>>> simgr = project.factory.simulation_manager()
>>> simgr.run()  # step until no more active states
Program running with argc=<SAO <BV64 0x0>> and argv=<SAO <BV64 0x7fffffffffeffa0>>
<SimulationManager with 1 deadended>
```

Now, whenever program execution reaches the main function, instead of executing the actual main function, it will execute this procedure! It just prints out a message, and returns.

Now, let's talk about what happens on the edge of this function! When entering the function, where do the values that go into the arguments come from? You can define your `run()` function with however many arguments you like, and the SimProcedure runtime will automatically extract from the program state those arguments for you, via a calling convention, and call your run function with them. Similarly, when you return a value from the run function, it is placed into the state (again, according to the calling convention), and the actual control-flow action of returning from a function is performed, which depending on the architecture may involve jumping to the link register or jumping to the result of a stack pop.

It should be clear at this point that the SimProcedure we just wrote is meant to totally replace whatever function it is hooked over top of. In fact, the original use case for SimProcedures was replacing library functions. More on that later.

Implementation Context

On a `Project` class, the dict `project._sim_procedures` is a mapping from address to `SimProcedure` instances. When the execution pipeline reaches an address that is present in that dict, that is, an address that is hooked, it will execute `project._sim_procedures[address].execute(state)`. This will consult the calling convention to extract the arguments, make a copy of itself in order to preserve thread safety, and run the `run()` method. It is important to produce a new instance of the SimProcedure for each time it is run, since the process of running a SimProcedure necessarily involves mutating state on the SimProcedure instance, so we need separate ones for each step, lest we run into race conditions in multithreaded environments.

kwargs

This hierarchy implies that you might want to reuse a single SimProcedure in multiple hooks. What if you want to hook the same SimProcedure in several places, but tweaked slightly each time? angr's support for this is that any additional keyword arguments you pass to the constructor of your SimProcedure will end up getting passed as keyword args to your SimProcedure's `run()` method. Pretty cool!

Data Types

If you were paying attention to the example earlier, you noticed that when we printed out the arguments to the `run()` function, they came out as a weird `<SAO <BV64 0xSTUFF>>` class. This is a `SimActionObject`. Basically, you don't need to worry about it too much, it's just a thin wrapper over a normal bitvector. It does a bit of tracking of what exactly you do with it inside the SimProcedure---this is helpful for static analysis.

You may also have noticed that we directly returned the python int `0` from the procedure. This will automatically be promoted to a word-sized bitvector! You can return a native number, a bitvector, or a SimActionObject.

When you want to write a procedure that deals with floating point numbers, you will need to specify the calling convention manually. It's not too hard, just provide a cc to the hook: `cc = project.factory.cc_from_arg_kinds((True, True), ret_fp=True)` and `project.hook(address, ProcedureClass(cc=mycc))` This method for passing in a calling convention works for all calling conventions, so if angr's autodetected one isn't right, you can fix that.

Control Flow

How can you exit a SimProcedure? We've already gone over the simplest way to do this, returning a value from `run()`. This is actually shorthand for calling `self.ret(value)`. `self.ret()` is the function which knows how to perform the specific action of returning from a function.

SimProcedures can use lots of different functions like this!

- `ret(expr)`: Return from a function
- `jump(addr)`: Jump to an address in the binary
- `exit(code)`: Terminate the program
- `call(addr, args, continue_at)`: Call a function in the binary
- `inline_call(procedure, *args)`: Call another SimProcedure in-line and return the results

That second-last one deserves some looking-at. We'll get there after a quick detour...

Conditional Exits

What if we want to add a conditional branch out of a SimProcedure? In order to do that, you'll need to work directly with the SimSuccessors object for the current execution step.

The interface for this is `self.successors.add_successor(state, addr, guard, jumpkind)`. All of these parameters should have an obvious meaning if you've followed along so far. Keep in mind that the state you pass in will NOT be copied and WILL be mutated, so be sure to make a copy beforehand if there will be more work to do!

SimProcedure Continuations

How can we call a function in the binary and have execution resume within our SimProcedure? There is a whole bunch of infrastructure called the "SimProcedure Continuation" that will let you do this. When you use `self.call(addr, args, continue_at)`, `addr` is expected to be the address you'd like to call, `args` is the tuple of arguments you'd like to call it with, and `continue_at` is the name of another method in your SimProcedure class that you'd like execution to continue at when it returns. This method must have the same signature as the `run()` method. Furthermore, you can pass the keyword argument `cc` as the calling convention that ought to be used to communicate with the callee.

When you do this, you finish your current step, and execution will start again at the next step at the function you've specified. When that function returns, it has to return to some concrete address! That address is specified by the SimProcedure runtime: an address is allocated in angr's externs segment to be used as the return site for returning to the given method call. It is then hooked with a copy of the procedure instance tweaked to run the specified `continue_at` function instead of `run()`, with the same args and kwargs as the first time.

There are two pieces of metadata you need to attach to your SimProcedure class in order to use the continuation subsystem correctly:

- Set the class variable `IS_FUNCTION = True`
- Set the class variable `local_vars` to a tuple of strings, where each string is the name of an instance variable on your SimProcedure whose value you would like to persist to when you return.
  Local variables can be any type so long as you don't mutate their instances.

You may have guessed by now that there exists some sort of auxiliary storage in order to hold on to all this data. You would be right! The state plugin `state.callstack` has an entry called `.procedure_data` which is used by the SimProcedure runtime to store information local to the current call frame. angr tracks the stack pointer in order to make the current top of the `state.callstack` a meaningful local data store. It's stuff that ought to be stored in memory in a stack frame, but the data can't be serialized and/or memory allocation is hard.

As an example, let's look at the SimProcedure that angr uses internally to run all the shared library initializers for a `full_init_state` for a linux program:

```
class LinuxLoader(angr.SimProcedure):
    NO_RET = True
    IS_FUNCTION = True
    local_vars = ('initializers',)

    def run(self):
        self.initializers = self.project.loader.initializers
        self.run_initializer()

    def run_initializer(self):
        if len(self.initializers) == 0:
            self.project._simos.set_entry_register_values(self.state)
            self.jump(self.project.entry)
        else:
            addr = self.initializers[0]
            self.initializers = self.initializers[1:]
            self.call(addr, (self.state.posix.argc, self.state.posix.argv, self.state.posi
```

This is a particularly clever usage of the SimProcedure continuations. First, notice that the current project is available for use on the procedure instance. This is some powerful stuff you can get yourself into; for safety you generally only want to use the project as a read-only or append-only data structure. Here we're just getting the list of dynamic intializers from the loader. Then, for as long as the list isn't empty, we pop a single function pointer out of the list, being careful not to mutate the list, since the list object is shared across states, and then call it, returning to the `run_initializer` function again. When we run out of initializers, we set up the entry state and jump to the program entry point.

Very cool!

Global Variables

As a brief aside, you can store global variables in `state.globals`. This is a dictionary that just gets shallow-copied from state to successor state. Because it's only a shallow copy, its members are the same instances, so the same rules as local variables in SimProcedure continuations apply. You need to be careful not to mutate any item that is used as a global variable unless you know exactly what you're doing.

Helping out static analysis

We've already looked at the class variable `IS_FUNCTION`, which allows you to use the SimProcedure continuation. There are a few more class variables you can set, though these ones have no direct benefit to you - they merely mark attributes of your function so that static analysis knows what it's doing.

- `NO_RET`: Set this to true if control flow will never return from this function
- `ADDS_EXITS`: Set this to true if you do any control flow other than returning
- `IS_SYSCALL`: Self-explanatory

Furthermore, if you set `ADDS_EXITS`, you may also want to define the method `static_exits()`. This function takes a single parameter, a list of IRSBs that would be executed in the run-up to your function, and asks you to return a list of all the exits that you know would be produced by your function in that case. The return value is expected to be a list of tuples of (address (int), jumpkind (str)). This is meant to be a quick, best-effort analysis, and you shouldn't try to do anything crazy or intensive to get your answer.

User Hooks

The process of writing and using a SimProcedure makes a lot of assumptions that you want to hook over a whole function. What if you don't? There's an alternate interface for hooking, a *user hook*, that lets you streamline the process of hooking sections of code.

```
>>> @project.hook(0x1234, length=5)
... def set_rax(state):
...     state.regs.rax = 1
```

This is a lot simpler! The idea is to use a single function instead of an entire SimProcedure subclass. No extraction of arguments is performed, no complex control flow happens.

Control flow is controlled by the length argument. After the function finishes executing in this example, the next step will start at 5 bytes after the hooked address. If the length argument is omitted or set to zero, execution will resume executing the binary code at exactly the hooked address, without re-triggering the hook. The `Ijk_NoHook` jumpkind allows this to happen.

If you want more control over control flow coming out of a user hook, you can return a list of successor states. Each successor will be expected to have `state.regs.ip`, `state.scratch.guard`, and `state.scratch.jumpkind` set. The IP is the target instruction pointer, the guard is a symbolic boolean representing a constraint to add to the state related to it being taken as opposed to the others, and the jumpkind is a VEX enum string, like `Ijk_Boring`, representing the nature of the branch.

The general rule is, if you want your SimProcedure to either be able to extract function arguments or cause a program return, write a full SimProcedure class. Otherwise, use a user hook.

Hooking Symbols

As you should recall from the section on loading a binary, dynamically linked programs have a list of symbols that they must import from the libraries they have listed as dependencies, and angr will make sure, rain or shine, that every import symbol gets resolved by *some* address, whether it's a real implementaion of the function or just a dummy address hooked with a do-nothing stub. As a result, you can just use the `Project.hook_symbol` API to hook the address referred to by a symbol!

This means that you can replace library functions with your own code. For instance, to replace `rand()` with a function that always returns a consistent sequence of values:

```
>>> class NotVeryRand(SimProcedure):
```

```
...     def run(self, return_values=None):
...         rand_idx = self.state.globals.get('rand_idx', 0) % len(return_values)
...         out = return_values[rand_idx]
...         self.state.globals['rand_idx'] = rand_idx + 1
...         return out

>>> project.hook_symbol('rand', NotVeryRand(return_values=[413, 612, 1025, 1111]))
```

Now, whenever the program tries to call `rand()`, it'll return the integers from the `return_values` array in a loop.

Writing State Plugins

If you want to store some data on a state and have that information propagated from successor to successor, the easiest way to do this is with `state.globals`. However, this can become obnoxious with large amounts of interesting data, doesn't work at all for merging states, and isn't very object-oriented.

The solution to these problems is to write a *State Plugin* - an appendix to the state that holds data and implements an interface for dealing with the lifecycle of a state.

My First Plugin

Let's get started! All state plugins are implemented as subclasses of `angr.SimStatePlugin`. Once you've read this document, you can use the API reference for this class to quickly review the semantics of all the interfaces you should implement.

The most important method you need to implement is `copy`: it should be annotated with the `memo` staticmethod and take a dict called the "memo"---these'll be important later---and returns a copy of the plugin. Short of that, you can do whatever you want. Just make sure to call the superclass initializer!

```
>>> import angr
>>> class MyFirstPlugin(angr.SimStatePlugin):
...     def __init__(self, foo):
...         super(MyFirstPlugin, self).__init__()
...         self.foo = foo
...
...     @angr.SimStatePlugin.memo
...     def copy(self, memo):
...         return MyFirstPlugin(self.foo)

>>> state = angr.SimState(arch='AMD64')
>>> state.register_plugin('my_plugin', MyFirstPlugin('bar'))
>>> assert state.my_plugin.foo == 'bar'

>>> state2 = state.copy()
>>> state.my_plugin.foo = 'baz'
>>> state3 = state.copy()
>>> assert state2.my_plugin.foo == 'bar'
>>> assert state3.my_plugin.foo == 'baz'
```

It works! Note that plugins automatically become available as attributes on the state. `state.get_plugin(name)` is also available as a more programmatic interface.

Where's the state?

State plugins have access to the state, right? So why isn't it part of the initializer? It turns out, there are a plethora of issues related to initialization order and dependency issues, so to simplify things as much as possible, the state is not part of the initializer but is rather set onto the state in a separate phase, by using the `set_state` method. You can override this state if you need to do things like propagate the state to subcomponents or extract architectural information.

```
>>> def set_state(self, state):
...     super(SimStatePlugin, self).set_state(state)
...     self.symbolic_word = claripy.BVS('my_variable', self.state.arch.bits)
```

Note the `self.state`! That's what the super `set_state` sets up.

However, there's no guarantee on what order the states will be set onto the plugins in, so if you need to interact with *other plugins* for initialization, you need to override the `init_state` method.

Once again, there's no guarantee on what order these will be called in, so the rule is to make sure you set yourself up good enough during `set_state` so that if someone else tries to interact with you, no type errors will happen. Here's an example of a good use of `init_state`, to map a memory region in the state. The use of an instance variable (presumably copied as part of `copy()`) ensures this only happens the first time the plugin is added to a state.

```
>>> def init_state(self):
...     if self.region is None:
...        self.region = self.state.memory.map_region(SOMEWHERE, 0x1000, 7)
```

Note: weak references

`self.state` is not the state itself, but rather a weak proxy to the state. You can still use this object as a normal state, but attempts to store it persistently will not work.

Merging

The other element besides copying in the state lifecycle is merging. As input you get the plugins to merge and a list of "merge conditions" - symbolic booleans that are the "guard conditions" describing when the values from each state should actually apply.

The important properties of the merge conditions are:

- They are mutually exclusive and span an entire domain - exactly one may be satisfied at once, and there will be additional constraints to ensure that at least one must be satisfied.
- `len(merge_conditions)` == len(others) + 1, since `self` counts too.
- `zip(merge_conditions, [self] + others)` will correctly pair merge conditions with plugins.

During the merge function, you should *mutate* `self` to become the merged version of itself and all the others, with respect to the merge conditions. This involves using the if-then-else structure that claripy provides. Here is an example of constructing this merged structure by merging a bitvector instance variable called `myvar`, producing a binary tree of if-then-else expressions searching for the correct condition:

```
for other_plugin, condition in zip(others, merge_conditions[1:]): # chop off self's condit
    self.myvar = claripy.If(condition, other_plugin.myvar, self.myvar)
```

This is such a common construction that we provide a utility to perform it automatically: `claripy.ite_cases`. The following code snippet is identical to the previous one:

```
self.myvar = claripy.ite_cases(zip(merge_conditions[1:], [o.myvar for o in others]), self.
```

Keep in mind that like the rest of the top-level claripy functions, `ite_cases` and `If` are also available from `state.solver`, and these versions will perform SimActionObject unwrapping if applicable.

Common Ancestor

The full prototype of the `merge` interface is `def merge(self, others, merge_conditions, common_ancestor=None)`. `others` and `merge_conditions` have been discussed in depth already.

The common ancestor is the instance of the plugin from the most recent common ancestor of the states being merged. It may not be available for all merges, in which case it will be None. There are no rules for how exactly you should use this to improve the quality of your merges, but you may find it useful in more complex setups.

Widening

There is another kind of merging called *widening* which takes several states and produces a more general state. It is used during static analysis.

TODO: @FISH PLEASE EXPLAIN WHAT THIS MEANS

Serialization

In order to support serialization of states which contain your plugin, you should implement the `__getstate__`/`__setstate__` magic method pair. Keep in mind the following guidelines:

- Your serialization result should *not* include the state.
- After deserialization, `set_state()` will be called again.

This means that plugins are "detached" from the state and serialized in an isolated environment, and then reattached to the state on deserialization.

Plugins all the way down

You may have components within your state plugins which are large and complicated and start breaking object-orientation in order to make copy/merge work well with the state lifecycle. You're in luck! Things can be state plugins even if they aren't directly attached to a state. A great example of this is `SimFile`, which is a state plugin but is stored in the filesystem plugin, and is never used with `SimState.register_plugin`. When you're doing this, there are a handful of rules to remember which will keep your plugins safe and happy:

- Annotate your copy function with `@SimStatePlugin.memo`.
- In order to prevent *divergence* while copying multiple references to the same plugin, make sure you're passing the memo (the argument to copy) to the `.copy` of any subplugins. This with the previous point will preserve object identity.
- In order to prevent *duplicate merging* while merging multiple references to the same plugin, there should be a concept of the "owner" of each instance, and only the owner should run the merge routine.
- While passing arguments down into sub-plugins `merge()` routines, make sure you unwrap `others` and `common_ancestor` into the appropriate types. For example, if `PluginA` contains a `PluginB`, the former should do the following:

```
>>> def merge(self, others, merge_conditions, common_ancestor=None):
...     # ... merge self
...     self.plugin_b.merge([o.plugin_b for o in others], merge_conditions,
...         common_ancestor=None if common_ancestor is None else common_ancestor.plugin_b)
```

Setting Defaults

To make it so that a plugin will automatically become available on a state when requested, without having to register it with the state first, you can register it as a *default*. The following code example will make it so that whenever you access `state.my_plugin`, a new instance of `MyPlugin` will be instanciated and registered with the state.

```
MyPlugin.register_default('my_plugin')
```

Extending the Environment Model

One of the biggest issues you may encounter while using angr to analyze programs is an incomplete model of the environment, or the APIs, surrounding your program. This usually takes the form of syscalls or dynamic library calls, or in rare cases, loader artifacts. angr provides a convenient interface to do most of these things!

Everything discussed here involves writing SimProcedures, so make sure you know how to do that!.

Note that this page should be treated as a narrative document, not a reference document, so you should read it at least once start to end.

Setup

You *probably* want to have a development install of angr, i.e. set up with the script in the angr-dev repository. It is remarkably easy to add new API models by just implementing them in certain folders of the angr repository. This is also desirable because any work you do in this field will almost always be useful to other people, and this makes it extremely easy to submit a pull request.

However, if you want to do your development out-of-tree, you want to work against a production version of angr, or you want to make customized versions of already-implemented API functions, there are ways to incorporate your extensions programmatically. Both these techniques, in-tree and out-of-tree, will be documented at each step.

Dynamic library functions - import dependencies

This is the easiest case, and the case that SimProcedures were originally designed for.

First, you need to write a SimProcedure representing the function. Then you need to let angr know about it.

Case 1, in-tree development: SimLibraries and catalogues

angr has a magical folder in its repository, angr/procedures. Within it are all the SimProcedure implementations that come bundled with angr as well as information about what libraries implement what functions.

Each folder in the `procedures` directory corresponds to some sort of *standard*, or a body that specifies the interface part of an API and its semantics. We call each folder a *catalog* of procedures. For example, we have `libc` which contains the functions defined by the C standard library, and a separate folder `posix` which contains the functions defined by the posix standard. There is some magic which automatically scrapes these folders in the `procedures` directory and organizes them into the

`angr.SIM_PROCEDURES` dict. For example, `angr/procedures/libc/printf.py` contains both `class printf` and `class __printf_chk`, so there exists both `angr.SIM_PROCEDURES['libc']['printf']` and `angr.SIM_PROCEDURES['libc']['__printf_chk']`.
The purpose of this categorization is to enable easy sharing of procedures among different libraries. For example. libc.so.6 contains all the C standard library functions, but so does msvcrt.dll! These relationships are represented with objects called `SimLibraries` which represent an actual shared library file, its functions, and their metadata. Take a look at the API reference for SimLibrary along with the code for setting up glibc to learn how to use it.

SimLibraries are defined in a special folder in the procedures directory, `procedures/definitions`. Files in here should contain an *instance*, not a subclass, of `SimLibrary`. The same magic that scrapes up SimProcedures will also scrape up SimLibraries and put them in `angr.SIM_LIBRARIES`, keyed on each of their common names. For example, `angr/procedures/definitions/linux_loader.py` contains `lib = SimLibrary(); lib.set_library_names('ld.so', 'ld-linux.so', 'ld.so.2', 'ld-linux.so.2', 'ld-linux-x86_64.so.2')`, so you can access it via `angr.SIM_LIBRARIES['ld.so']` or `angr.SIM_LIBRARIES['ld-linux.so']` or any of the other names.

At load time, all the dynamic library dependencies are looked up in `SIM_LIBRARIES` and their procedures (or stubs!) are hooked into the project's address space to summarize any functions it can. The code for this process is found here.

SO, the bottom line is that you can just write your own SimProcedure and SimLibrary definitions, drop them into the directory structure, and they'll automatically be applied. If you're adding a procedure to an existing library, you can just drop it into the appropriate catalog and it'll be picked up by all the libraries using that catalog, since most libraries construct their list of function implementation by batch-adding entire catalogs.

Case 2, out-of-tree development, tight integration

If you'd like to implement your procedures outside the angr repository, you can do that. You effectively do this by just manually adding your procedures to the appropriate SimLibrary. Just call `angr.SIM_LIBRARIES[libname].add(name, proc_cls)` to do the registration.

Note that this will only work if you do this before the project is loaded with `angr.Project`. Note also that adding the procedure to `angr.SIM_PROCEDURES`, i.e. adding it directly to a catalog, will *not* work, since these catalogs are used to construct the SimLibraries only at import and are used by value, not by reference.

Case 3, out-of-tree development, loose integration

Finally, if you don't want to mess with SimLibraries at all, you can do things purely on the project level with `hook_symbol`.

Syscalls

Unlike dynamic library methods, syscall procedures aren't incorporated into the project via hooks. Instead, whenever a syscall instruction is encountered, the basic block should end with a jumpkind of `Ijk_Sys`. This will cause the next step to be handled by the SimOS associated with the project, which will extract the syscall number from the state and query a specialized SimLibrary with that.

This deserves some explanation.

There is a subclass of SimLibrary called SimSyscallLibrary which is used for collecting all the functions that are part of an operating system's syscall interface. SimSyscallLibrary uses the same system for managing implementations and metadata as SimLibrary, but adds on top of it a system for managing syscall numbers for multiple ABIs (application binary interfaces, like an API but lower level). The best example for an implementation of a SimSyscallLibrary is the linux syscalls. It keeps its procedures in a normal SimProcedure catalog called `linux_kernel` and adds them to the library, then adds several syscall number mappings, including separate mappings for `mips-o32`, `mips-n32`, and `mips-n64`.

In order for syscalls to be supported in the first place, the project's SimOS must inherit from `SimUserland`, itself a SimOS subclass. This requires the class to call SimUserland's constructor with a super() call that includes the `syscall_library` keyword argument, specifying the specific SimSyscallLibrary that contains the appropriate procedures and mappings for the operating system. Additionally, the class's `configure_project` must perform a super() call including the `abi_list` keyword argument, which contains the list of ABIs that are valid for the current architecture. If the ABI for the syscall can't be determined by just the syscall number, for example, that amd64 linux programs can use either `int 0x80` or `syscall` to invoke a syscall and these two ABIs use overlapping numbers, the SimOS cal override `syscall_abi()`, which takes a SimState and returns the name of the current syscall ABI. This is determined for int80/syscall by examining the most recent jumpkind, since libVEX will produce different syscall jumpkinds for the different instructions.

Calling conventions for syscalls are a little weird right now and they ought to be refactored. The current situation requires that `angr.SYSCALL_CC` be a map of maps `{arch_name: {os_name: cc_cls}}`, where `os_name` is the value of project.simos.name, and each of the calling convention classes must include an extra method called `syscall_number` which takes a state and return the current syscall number. Look at the bottom of `calling_conventions.py` to learn more about it. Not very object-oriented at all...

As a side note, each syscall is given a unique address in a special object in CLE called the "kernel object". Upon a syscall, the address for the specific syscall is set into the state's instruction pointer, so it will show up in the logs. These addresses are not hooked, they are just used to identify syscalls during analysis given only an address trace. The test for determining if an address corresponds to a syscall is `project.simos.is_syscall_addr(addr)` and the syscall corresponding to the address can be retrieved with `project.simos.syscall_from_addr(addr)`.

Case 1, in-tree development

SimSyscallLibraries are stored in the same place as the normal SimLibraries, `angr/procedures/definitions`. These libraries don't have to specify any common name, but they can if they'd like to show up in `SIM_LIBRARIES` for easy access.

The same thing about adding procedures to existing catalogs of dynamic library functions also applies to syscalls - implementing a linux syscall is as easy as writing the SimProcedure and dropping the implemementation into `angr/procedures/linux_kernel`. As long as the class name matches one of the names in the number-to-name mapping of the SimLibrary (all the linux syscall numbers are included with

To add a new operating system entirely, you need to implement the SimOS as well, as a subclass of SimUserland. To integrate it into the tree, you should add it to the `simos` directory, but this is not a magic directory like `procedures`. Instead, you should add a line to `angr/simos/__init__.py` calling `register_simos()` with the OS name as it appears in `project.loader.main_object.os` and the SimOS class. Your class should do everything described above.

Case 2, out-of-tree development, tight integration

You can add syscalls to a SimSyscallLibrary the same way you can add functions to a normal SimLibrary, by tweaking the entries in `angr.SIM_LIBRARIES`. If you're this for linux you want `angr.SIM_LIBRARIES['linux'].add(name, proc_cls)`.

You can register a SimOS with angr from out-of-tree as well - the same `register_simos` method is just sitting there waiting for you as `angr.simos.register_simos(name, simos_cls)`.

Case 3, out-of-tree development, loose integration

The SimSyscallLibrary the SimOS uses is copied from the original during setup, so it is safe to mutate. You can directly fiddle with `project.simos.syscall_library` to manipulate an individual project's syscalls.

You can provide a SimOS class (not an instance) directly to the `Project` constructor via the `simos` keyword argument, so you can specify the SimOS for a project explicitly if you like.

SimData

What about when there is an import dependency on a data object? This is easily resolved when the given library is actually loaded into memory - the relocation can just be resolved as normal. However, when the library is not loaded (for example, `auto_load_libs=False`, or perhaps some dependency is simply missing), things get tricky. It is not possible to guess in most cases what the value should be, or even what its size should be, so if the guest program ever dereferences a pointer to such a symbol, emulation will go off the rails.

CLE will warn you when this might happen:

```
[22:26:58] [cle.backends.externs] |  WARNING: Symbol was allocated without a known size; e
[22:26:58] [cle.backends.externs] |  WARNING: Symbol was allocated without a known size; e

[22:26:58] [cle.backends.externs] |  WARNING: Symbol was allocated without a known size; e
[22:26:58] [cle.backends.externs] |  WARNING: Symbol was allocated without a known size; e
```

If you see this message and suspect it is causing issues (i.e. the program is actually introspecting the value of these symbols), you can resolve it by implementing and registering a SimData class, which is like a SimProcedure but for data. Simulated data. Very cool.

A SimData can effectively specify some data that must be used to provide an unresolved import symbol. It has a number of mechanisms to make this more useful, including the ability to specify relocations and subdependencies.
Look at the SimData class reference and the existing SimData subclasses for guidelines on how to do this.

TODO: Writing Exploration Techniques

Writing Analyses

An analysis can be created by subclassing the `angr.Analysis` class. In this section, we'll create a mock analysis to show off the various features. Let's start with something simple:

```
>>> import angr

>>> class MockAnalysis(angr.Analysis):
...     def __init__(self, option):
...         self.option = option

>>> angr.AnalysesHub.register_default('MockAnalysis', MockAnalysis) # register the class wi
```

This is a very simple analysis -- it takes an option, and stores it. Of course, it's not useful, but this is just a demonstration.

Let's see how to run our new analysis:

```
>>> proj = angr.Project("/bin/true")
>>> mock = proj.analyses.MockAnalysis('this is my option')
>>> assert mock.option == 'this is my option'
```

Working with projects

Via some python magic, your analysis will automatically have the project upon which you are running it under the `self.project` property. Use this to interact with your project and analyze it!

```
>>> class ProjectSummary(angr.Analysis):
...     def __init__(self):
...         self.result = 'This project is a %s binary with an entry point at %#x.' % (sel

>>> angr.AnalysesHub.register_default('ProjectSummary', ProjectSummary)
>>> proj = angr.Project("/bin/true")

>>> summary = proj.analyses.ProjectSummary()
>>> print(summary.result)
This project is a AMD64 binary with an entry point at 0x401410.
```

Analysis Resilience

Sometimes, your (or our) code might suck and analyses might throw exceptions. We understand, and we also understand that oftentimes a partial result is better than nothing. This is specifically true when, for example, running an analysis on all of the functions in a program. Even if some of the functions fails, we still want to know the results of the functions that do not.

To facilitate this, the `Analysis` base class provides a resilience context manager under `self._resilience`. Here's an example:

```
>>> class ComplexFunctionAnalysis(angr.Analysis):
...     def __init__(self):
...         self._cfg = self.project.analyses.CFG()
...         self.results = { }
...         for addr, func in self._cfg.function_manager.functions.items():
...             with self._resilience():
...                 if addr % 2 == 0:
...                     raise ValueError("can't handle functions at even addresses")
...                 else:
...                     self.results[addr] = "GOOD"
```

The context manager catches any exceptions thrown and logs them (as a tuple of the exception type, message, and traceback) to `self.errors`. These are also saved and loaded when the analysis is saved and loaded (although the traceback is discarded, as it is not picklable).

You can tune the effects of the resilience with two optional keyword parameters to `self._resilience()`.

The first is `name`, which affects where the error is logged. By default, errors are placed in `self.errors`, but if `name` is provided, then instead the error is logged to `self.named_errors`, which is a dict mapping `name` to a list of all the errors that were caught under that name. This allows you to easily tell where thrown without examining its traceback.

The second argument is `exception`, which should be the type of the exception that `_resilience` should catch. This defaults to `Exception`, which handles (and logs) almost anything that could go wrong. You can also pass a tuple of exception types to this option, in which case all of them will be caught.

Using `_resilience` has a few advantages:

1. Your exceptions are gracefully logged and easily accessible afterwards. This is really nice for writing testcases.
2.

When creating your analysis, the user can pass `fail_fast=True`, which transparently disable the resilience, which is really nice for manual testing.

3. It's prettier than having `try`/`except` everywhere.

Have fun with analyses! Once you master the rest of angr, you can use analyses to understand anything computable!

TODO: Adding Support for New Architectures

Scripting angr management

Please note that the documentation and the API for angr management are highly in-flux. You will need to spend time reading the source code. Grep is your friend. If you have questions, please ask in the angr slack.

If you build something which uses an API and you want to make sure it doesn't break, you can contribute a testcase for the API!

This codebase is absolutely filled to the brim with one-off hacks. If you see some code and think, "hm, that doesn't seem like an extensible or best-practices way to code that", you're probably right. Cleaning up angr management's code is a top priority for us, so if you have some ideas to fix these sorts of issues, please let us know, either in an issue or a pull request!

The console, and the basic objects

angr management opens with an IPython console ready for input. This console has in its namespace several objects which are important for manipulating angr management and its data.

- First, the `main_window`. This is the `QMainWindow` instance for the application. It contains basic functions that correspond to top-level buttons, such as loading a binary.
- Next, the `workspace`. This is a light object which coordinates the UI elements and manages the tabbed environment. You can use it to access any analysis-related GUI element, such as the disassembly view.
- Finally, the `instance`. This is angr management's data model. It contains mechanisms for synchronizing components on shared data sources, as well as logic for creating long-running jobs.

`workspace` is also available as an attribute on `main_window` and `instance` is available as an attribute on `workspace`. If you are programming in a namespace where none of these objects are available, you can import the `angrmanagment.logic.GlobalInfo` object, which contains a reference to `main_window`.

The ObjectContainer

angr management uses a class called ObjectContainer to implement a pub-sub model and synchronize changing object references. Let's use `instance.project` as an example. This is an ObjectContainer that contains the current project. You can use it in every way that you would normally use a project - you can access `project.factory`, `project.kb`, etc. However, it also has two very important features that are helpful for building UIs.

First, the pub-sub model. You can subscribe to changes to this object by calling `instance.project.am_subscribe(callback)`. Then, you can notify listeners of changes by calling `instance.project.am_event()`. Note that events are NEVER automatically triggered - you must call `am_event` in order to trigger the callbacks. One useful feature of this model is that you can provide arbitrary keyword arguments to `am_event`, and they will be passed on to each callback. This means that you should always have your callbacks take `**kwargs` in order to account for unknown parameters. This feature is particularly useful to prevent feedback loops - if you ever find yourself in a situation where you need to broadcast an event from your callback, you can add an argument that you can use as a flag not to recurse any further.

Next, object reference mutability. Let's say you have a widget that displays information about the project. Following the principle of least access, you should only provide as much information as is necessary to do the job - in this case, just the project object. If you provide the basic project object, this will cause issues when a new project is loaded. Notably, there will be a dangling reference held to the original project, preventing it from being garbage collected, and the widget will not update, continuing to show the old project's information. Now, if you provide the project's ObjectContainer, a new project can be created and inserted into the container and the reference will instantly be available to your widget. If you ever wanted to load a new project yourself, all you have to do is assign to `instance.project.am_obj` and then send off an event. Combined with the event publication model, this provides an efficient way to build responsive UIs that follow the principle of least access.

One important way that you can't use the object container the same way that you would a normal object is that `is None` will obviously not work. To resolve this, you can use `instance.project.am_none` - this will be True when no project is loaded.

One interesting feature of the ObjectContainer is that they can nest. If you have a container which contains a container which contains an object, any events sent to the inner container will also be sent to subscribers to the outer container. This allows patterns such as the list of SimStates actually containing a list of ObjectContainers which contain states, and the "current state" container actually contains one of these containers. The result of this is that UI elements can either subscribe to the current state, no matter

A full list of standard ObjectContainers that can be found in the instance `__init__` method. There are more containers floating around for synchronizing on non-global elements - for example, the current state of the disassembly view is synchronized through its InfoDock object. Given a disassembly view instance, you can subscribe to, for example, its current selected instructions through `view.infodock.selected_insns`.

Manipulating UI elements

The `workspace` contains methods to manipulate UI elements. Notably, you can manipulate all open tabs with the `workspace.view_manager` reference. Additionally, you can pass any sort of object you like to `workspace.viz()` and it will attempt to visualize the object in the current window.

Writing plugins

angr management has a very flexible plugin framework. A plugin is a python file containing a subclass of `angrmanagement.plugins.BasePlugin`. Plugin files will be automatically loaded from the `plugins` module of angr management, and also from `~/.local/share/angr-management/plugins`. These paths are configurable through the program configuration, but at the time of writing, this is not exposed in the UI.

The best way to see the tools you can use while building a plugin is to read the plugin base class source code. Any method or attribute can be overridden from a base class and will be automatically called on relevant events.

Writing tests

Look at the existing tests for examples. Generally, you can test UI components by creating the component and driving input to it via QTest. You can create a headless MainWindow instance by passing `show=False` to its constructor - this will also get you access to a workspace and an instance.

Examples

To help you get started with angr, we've created several examples. We've tried to organize them into major categories, and briefly summarize that each example will expose you to. Enjoy!

There are also a great amount of slightly more redundant examples (these mostly stem from CTF problems solved with angr by Shellphish) here.

If you want a high-level cheatsheet of the "techniques" used in the examples, see the angr strategies cheatsheet by Florent Bordignon.

To jump to a specific category:

- Introduction - examples showing off the very basics of angr's functionality
- Reversing - examples showing angr being used in reverse engineering tasks
- Vulnerability Discovery - examples of angr being used to search for vulnerabilities
- Exploitation - examples of angr being used as an exploitation assistance tool

---

Introduction

These are some introductory examples to give an idea of how to use angr's API.

Fauxware

This is a basic script that explains how to use angr to symbolically execute a program and produce concrete input satisfying certain conditions.

Binary, source, and script are found here.

Reversing

These are examples that use angr to solve reverse engineering challenges. There are a lot of these. We've chosen the most unique ones, and relegated the rest to the CTF Challenges section below.

Beginner reversing example: little_engine

```
Script author: Michael Reeves (github: @mastermjr)
Script runtime: 3 min 26 seconds (206 seconds)
Concepts presented:
stdin constraining, concrete optimization with Unicorn
```

This challenge is similar to the csaw challenge below, however the reversing is much more simple. The original code, solution, and writeup for the challenge can be found at the b01lers github here.

The angr solution script is here and the binary is here.

Whitehat CTF 2015 - Crypto 400

```
Script author: Yan Shoshitaishvili (github: @Zardus)
Script runtime: 30 seconds
Concepts presented: statically linked binary (manually hooking with function summaries), c
```

We solved this crackme with angr's help. The resulting script will help you understand how angr can be used for crackme *assistance*, not a full-out solve. Since angr cannot solve the actual crypto part of the challenge, we use it just to reduce the keyspace, and brute-force the rest.

You can find this script here and the binary here.

CSAW CTF 2015 Quals - Reversing 500, "wyvern"

```
Script author: Audrey Dutcher (github: @rhelmot)
Script runtime: 15 mins
Concepts presented: stdin constraining, concrete optimization with Unicorn
```

angr can outright solve this challenge with very little assistance from the user. The script to do so is here and the binary is here.

TUMCTF 2016 - zwiebel

```
Script author: Fish
Script runtime: 2 hours 31 minutes with pypy and Unicorn - expect much longer with CPython
Concepts presented: self-modifying code support, concrete optimization with Unicorn
```

This example is of a self-unpacking reversing challenge. This example shows how to enable Unicorn support and self-modification support in angr. Unicorn support is essential to solve this challenge within a reasonable amount of time - simulating the unpacking code symbolically is *very* slow. Thus, we execute it concretely in unicorn/qemu and only switch into symbolic execution when needed.

You may refer to other writeup about the internals of this binary. I didn’t reverse too much since I was pretty confident that angr is able to solve it :-)

The long-term goal of optimizing angr is to execute this script within 10 minutes. Pretty ambitious :P

Here is the binary and the script.

FlareOn 2015 - Challenge 5

```
Script author: Adrian Tang (github: @tangabc)
Script runtime: 2 mins 10 secs
Concepts presented: Windows support
```

This is another reversing challenge from the FlareOn challenges.

"The challenge is designed to teach you about PCAP file parsing and traffic decryption by reverse engineering an executable used to generate it. This is a typical scenario in our malware analysis practice where we need to figure out precisely what the malware was doing on the network"

For this challenge, the author used angr to represent the desired encoded output as a series of constraints for the SAT solver to solve for the input.

For a detailed write-up please visit the author's post here and you can also find the solution from the FireEye here

0ctf quals 2016 - trace

```
Script author: WGH (wgh@bushwhackers.ru)
Script runtime: 1 min 50 secs (CPython 2.7.10), 1 min 12 secs (PyPy 4.0.1)
Concepts presented: guided symbolic tracing
```

In this challenge we're given a text file with trace of a program execution. The file has two columns, address and instruction executed. So we know all the instructions being executed, and which branches were taken. But the initial data is not known.

Reversing reveals that a buffer on the stack is initialized with known constant string first, then an unknown string is appended to it (the flag), and finally it's sorted with some variant of quicksort. And we need to find the flag somehow.

angr easily solves this problem. We only have to direct it to the right direction at every branch, and the solver finds the flag at a glance.

Files are here.

ASIS CTF Finals 2015 - license

```
Script author: Fish Wang (github: @ltfish)
Script runtime: 3.6 sec
Concepts presented: using the filesystem, manual symbolic summary execution
```

This is a crackme challenge that reads a license file. Rather than hooking the read operations of the flag file, we actually pass in a filesystem with the correct file created.

Here is the binary and the script.

DEFCON Quals 2017 - Crackme2000

```
Script author: Shellphish
Script runtime: varies, but on the order of seconds
Concepts presented: automated reverse engineering
```

DEFCON Quals had a whole category for automatic reversing in 2017. Our scripts are here.

Vulnerability Discovery

These are examples of angr being used to identify vulnerabilities in binaries.

Beginner vulnerability discovery example: strcpy_find

```
Script author: Kyle Ossinger (github: @k0ss)
Concepts presented: exploration to vulnerability, programmatic find condition
```

This is the first in a series of "tutorial scripts" I'll be making which use angr to find exploitable conditions in binaries. The first example is a very simple program. The script finds a path from the main entry point to `strcpy`, but **only** when we control the source buffer of the `strcpy` operation. To hit the right path, angr has to solve for a password argument, but angr solved this in less than 2 seconds on my machine using the standard python interpreter. The script might look large, but that's only because I've heavily commented it to be more helpful to beginners. The challenge binary is here and the script is here.

CGC crash identification

```
Script author: Antonio Bianchi, Jacopo Corbetta
Concepts presented: exploration to vulnerability
```

This is a very easy binary containing a stack buffer overflow and an easter egg. CADET_00001 is one of the challenge released by DARPA for the Cyber Grand Challenge: link The binary can run in the DECREE VM:

link A copy of the original challenge and the angr solution is provided here CADET_00001.adapted (by

Grub "back to 28" bug

```
Script author: Audrey Dutcher (github: @rhelmot)
Concepts presented: unusal target (custom function hooking required), use of exploration t
```

This is the demonstration presented at 32c3. The script uses angr to discover the input to crash grub's password entry prompt.

script - vulnerable module

Exploitation

These are examples of angr's use as an exploitation assistance engine.

Insomnihack Simple AEG

```
Script author: Nick Stephens (github: @NickStephens)
Concepts presented: automatic exploit generation, global symbolic data tracking
```

Demonstration for Insomni'hack 2016. The script is a very simple implementation of AEG.

script

SecuInside 2016 Quals - mbrainfuzz - symbolic exploration for exploitability conditions

```
Script author: nsr (nsr@tasteless.eu)
Script runtime: ~15 seconds per binary
Concepts presented: symbolic exploration guided by static analysis, using the CFG
```

Originally, a binary was given to the ctf-player by the challenge-service, and an exploit had to be crafted automatically. Four sample binaries, obtained during the ctf, are included in the example. All binaries follow the same format; the command-line argument is validated in a bunch of functions, and when every check succeeds, a memcpy() resulting into a stack-based buffer overflow is executed. angr is used to find the way through the binary to the memcpy() and to generate valid inputs to every checking function individually.

The sample binaries and the script are located here and additional information be found at the author's Write-Up.

SECCON 2016 Quals - ropsynth

```
Script author: Yan Shoshitaishvili (github @zardus) and Nilo Redini
Script runtime: 2 minutes
```

This challenge required the automatic generation of ropchains, with the twist that every ropchain was succeeded by an input check that, if not passed, would terminate the application. We used symbolic execution to recover those checks, removed the checks from the binary, used angrop to build the ropchains, and instrumented them with the inputs to pass the checks.

The various challenge files are located here, with the actual solve script here.

Appendix

List of Claripy Operations

Arithmetic and Logic

Name	Description	Example
LShR	Logically shifts an expression to the right. (the default shifts are arithmetic)	`x.LShR(10)`
RotateLeft	Rotates an expression left	`x.RotateLeft(8)`
RotateRight	Rotates an expression right	`x.RotateRight(8)`
And	Logical And (on boolean expressions)	`solver.And(x == y, x 0)`
Or	Logical Or (on boolean expressions)	`solver.Or(x == y, y < 10)`
Not	Logical Not (on a boolean expression)	`solver.Not(x == y)` is th same as `x != y`
If	An If-then-else	Choose the maximum of two expressions: `solver.If(x y, x, y)`
ULE	Unsigned less than or equal to	Check if x is less than or equal to y: `x.ULE(y)`
ULT	Unsigned less than	Check if x is less than y: `x.ULT(y)`
UGE	Unsigned greater than or equal to	Check if x is greater than or equal to y: `x.UGE(y)`

UGT	Unsigned greater than	Check if x is greater than y: `x.UGT(y)`
SLE	Signed less than or equal to	Check if x is less than or equal to y: `x.SLE(y)`
SLT	Signed less than	Check if x is less than y: `x.SLT(y)`
SGE	Signed greater than or equal to	Check if x is greater than or equal to y: `x.SGE(y)`
SGT	Signed greater than	Check if x is greater than y: `x.SGT(y)`

TODO: Add the floating point ops

Bitvector Manipulation

Name	Description	Example
SignExt	Pad a bitvector on the left with `n` sign bits	`x.sign_extend(n)`
ZeroExt	Pad a bitvector on the left with `n` zero bits	`x.zero_extend(n)`
Extract	Extracts the given bits (zero-indexed from the *right*, inclusive) from an expression.	Extract the least significant byt of x: `x[7:0]`
Concat	Concatenates any number of expressions together into a new expression.	`x.concat(y, ...)`

Extra Functionality

There's a bunch of prepackaged behavior that you *could* implement by analyzing the ASTs and composing sets of operations, but here's an easier way to do it:

- You can chop a bitvector into a list of chunks of `n` bits with `val.chop(n)`
- You can endian-reverse a bitvector with `x.reversed`
- You can get the width of a bitvector in bits with `val.length`

- You can test if an AST has any symbolic components with `val.symbolic`
- You can get a set of the names of all the symbolic variables implicated in the construction of an AST with `val.variables`

List of State Options

State Modes

These may be enabled by passing `mode=xxx` to a state constructor.

Mode name	Description
`symbolic`	The default mode. Useful for most emulation and analysis tasks.
`symbolic_approximating`	Symbolic mode, but enables approximations for constraint solving.
`static`	A preset useful for static analysis. The memory model becomes an abstract region-mapping system, "fake return" successors skipping calls ar added, and more.
`fastpath`	A preset for extremely lightweight static analysis. Executing will skip all intensive processing to give a quick view of the behavior of code.
`tracing`	A preset for attempting to execute concretely through a program with a given input. Enables unicorn, enables resilience options, and will attempt to emulate access violations correctly.

Option Sets

These are sets of options, found as `angr.options.xxx`.

Set name	Description
`common_options`	Options necessary for basic execution
`symbolic`	Options necessary for basic symbolic execution
	Options that harden angr's emulation against

`resilience`	unsupported operations, attempting to carry on by treating the result as an unconstrained symbolic value and logging the occasion to `state.history.events`.
`refs`	Options that cause angr to keep a log of all the memory, register, and temporary references complete with dependency information in `history.actions`. This option consumes a lo of memory, so be careful!
`approximation`	Options that enable approximations of constraint solves via value-set analysis instead of calling int z3
`simplification`	Options that cause data to be run through z3's simplifiers before it reaches memory or register storage
`unicorn`	Options that enable the unicorn engine for executing on concrete data

Options

These are individual option objects, found as `angr.options.XXX`.

Option name	Description	Sets	Modes	Implicit adds
`ABSTRACT_MEMORY`	Use `SimAbstractMemory` to model memory as discrete regions		`static`	
`ABSTRACT_SOLVER`	Allow splitting constraint sets during simplification		`static`	
`ACTION_DEPS`	Track dependencies in SimActions			
`APPROXIMATE_GUARDS`	Use VSA when evaluating guard conditions			
`APPROXIMATE_`				

`MEMORY_INDICES`	Use VSA when evaluating memory indices	`approximation`	`symbolic_approximating`	
`APPROXIMATE_MEMORY_SIZES`	Use VSA when evaluating memory load/store sizes	`approximation`	`symbolic_approximating`	
`APPROXIMATE_SATISFIABILITY`	Use VSA when evaluating state satisfiability	`approximation`	`symbolic_approximating`	
`AST_DEPS`	Enables dependency tracking for all claripy ASTs			During executio
`AUTO_REFS`	An internal option used to track dependencies in SimProcedures			During executio
`AVOID_MULTIVALUED_READS`	Return a symbolic value without touching memory for any read that has a symbolic address		`fastpath`	
`AVOID_MULTIVALUED_WRITES`	Do not perfrom any write that has a symbolic address		`fastpath`	
`BEST_EFFORT_MEMORY_STORING`	Handle huge writes of symbolic size by pretending they are actually smaller		`static`, `fastpath`	
`BREAK_SIRSB_END`	Debug: trigger a breakpoint at the end of each block			
`BREAK_SIRSB_START`	Debug: trigger a breakpoint at the start of each block			
	Debug: trigger a			

`BREAK_SIRSTMT_END`	breakpoint at the end of each IR statement			
`BREAK_SIRSTMT_START`	Debug: trigger a breakpoint at the start of each IR statement			
`BYPASS_ERRORED_IRCCALL`	Treat clean helpers that fail with errors as returning unconstrained symbolic values	`resilience`	`fastpath`, `tracing`	
`BYPASS_ERRORED_IROP`	Treat operations that fail with errors as returning unconstrained symbolic values	`resilience`	`fastpath`, `tracing`	
`BYPASS_UNSUPPORTED_IRCCALL`	Treat unsupported clean helpers as returning unconstrained symbolic values	`resilience`	`fastpath`, `tracing`	

Changelog

This lists the *major* changes in angr. Tracking minor changes are left as an exercise for the reader :-)

angr 9.1

- (#2961) Refactored SimCC to support passing and returning structs and arrays by value
- (#2964) Functions from the knowledge base may now be pretty-printed, showing colors and reference arrows
- Improved `import angr` speed substantially
- (#2948) RDA's `dep_graph` can now be used to track dependencies between temporaries, constants, guard conditions, and function calls - if you want it!
- (#2929) Basic support for structs with bitfields in SimType
- There's a decompiler now

angr 9.0

- Switched to a new versioning scheme: major.minor.build_id

angr 8.19.7.25

- (#1503) Implement necessary helpers and information storage for call pretty printing
- (#1546) Add a new state option MEMORY_FIND_STRICT_SIZE_LIMIT
- (#1548) SimProcedure.static_exits: Allow providing name hints
- (cle#177) Use Enums for Symbol Types
- (cle#193) Add support for "named regions"
- (claripy#151) Implement operator precedence in claripy op rendering
- Added support for interaction recording in angr-management
- Several new simprocedure implementations
- Substantial imporvments to our CFG

angr 8.19.4.5

- (#1234) Massive improvements to CFG recovery for ARM and ARM cortex-m binaries.
- (#1416) Added support for analyzing Java programs via the Soot IR, including the ability to analyze interplay between Java code and JNI libraries. This branch was two years old!
- (#1427) Added a MemoryWatcher exploration technique to take action when the system is running out of RAM. Thanks @bannsec.
- (#1432) Added a `state.heap` plugin which manages the heap (with pluggable heap schemes!) and provides malloc functionality. Thanks @tgduckworth.
- Speed improvements for using the VEX engine and working with concrete data.
- Added SimLightRegisters, an alternate registers plugin that eliminates the abstraction of the register file for performance improvements at the cost of removing all instrumentability.
- `__version__` variable has been added to all modules.
- The `stack_base` kwarg for `call_state` is not broken for the first time ever
- https://github.com/python/cpython/pull/11384

angr 8.19.2.4

- (#1279) Support C++ function name demangling via itanium-demangler. Thanks @fmagin.
- (#1283) `_security_cookie` is initialized for SimWindows. Thanks @zeroSteiner.
- (#1298) Introduce `SimData`. It's a cleaner interface to deal with data imports in CLE -- especially for those data entries that are not imported because of missing or unloaded libraries. This commit fixes long-standing issues #151 and #693.
- (#1299, #1300, #1301, #1313, #1314, #1315, #1336, #1337, #1343, ...) Multiple CFGFast-related improvements and bug fixes.
- (#1332) `UnresolvableTarget` is now split into two classes: `UnresolvableJumpTarget` and `UnresolvableCallTarget`. Thanks @Kyle-Kyle.
- (#1382) Add a preliminary implementation of angr decompiler. Give it a try! `p = angr.Project("cfg_loop_unrolling", auto_load_libs=False); p.analyses.CFG(); print(p.analyses.Decompiler(p.kb.functions['test_func']).codegen.text)`.
- (#1421) `SimAction`s now have incrementing IDs. Thanks @bannsec.
- (#1408) `ANA`, angr's old identity-aware serialization backend, has been removed. Instead of non-obvious serialization behavior, all angr objects should now be pickleable. If one is not, please file an issue. For use-cases that require identity-awareness (i.e., deduplicating ASTs across states serialized at different times), an `angr.vaults` module has been introduced.
- Added a facility to synchronize state between angr and a running target a la avatar2
- Changed unconstrained registers/memory warning to be less obnoxious and contain useful information. Also added `SYMBOL_FILL_UNCONSTRAINED_REGISTERS` and `SYMBOL_FILL_UNCONSTRAINED_MEMORY` state options to silence them.

angr 8.18.10.25

- The IDA backend for CLE has been removed. It has been broken for quite some time, but now it has been disabled for your own safety.
- Surveyors have been removed! Finally! This is thanks to @danse-macabre who contributed an Exploration Technique for the Slicecutor. Backwards slicing has now been brought out of the angr dark ages.
- SimCC can now be initialized with a string containing C function prototype in its `func_ty` argument
- Similarly, Callable can now be run with its arguments instanciated from a string containing C expressions
- Tracer has been substantially refactored - it will now handle more kinds of desyncs, ASLR slides, and is much more friendly for hacking. We will be continuing to improve it!
- The Oppologist and Driller have been refactored to play nice with other exploration techniques
- SimProcedure continuations now have symbols in the externs object, so `describe_addr` will work on them. Additionally, the representation for SimProcedure (appearing in `history.descriptions`

and `project._sim_procedures` among other places) has been improved to show this information.

angr 8.18.10.5

Largely a bugfix release, but with a few bonus treats:

- API documentation has been rewritten for Exploration Technique. It should be much easier to use now.
- Simulation Manager will throw an error if you pass incorrect keyword arguments (??? why was it like this)
- The `save_unconstrained` flag of Simulation Manager is now on by default
- If a step produces only unsatisfiable states, they will appear in the `'unsat'` stash regardless of the `save_unsat` setting, since this usually indicates a bug. Add `unsat` to the `auto_drop` parameter to restore the old behavior.

angr 8.18.10.1

Welcome to angr 8! The biggest change for this major version bump is the transition to python 3. You can read about this, as well as a few other breaking changes, in the migration guide.

- Switch to python 3
- Refactor to Clemory to clean up the API and speed things up drastically
- Remove `object.symbols_by_addr` (dict) and add `object.symbols` (sorted list); add `fuzzy` parameter to `loader.find_symbol`
- CFGFast is much, much faster now. CFGAccurate has been renamed to CFGEmulated.
- Support for avx2 unpack instructions, courtesy of D. J. Bernstein
- Removed support for immutable simulation managers
- angr will now show you a warning when using uninitialized memory or registers
- angr will now NOT show you a warning if you have a capstone 3.x install unless you're actually interacting with the relevant missing parts
- Many, many, many bug fixes

angr 7.8.7.1

- Remove `LoopLimiter` and `DFG`.
- (#1063) `CFGAccurate` can now leverage indirect jump resolvers to resolve indirect jumps.

angr 7.8.6.23

- (PyVEX!#134) We now recognize LDMDB r11, {xxx, pc} as a ret instruction for ARM.
- (#1053) CFGFast spends less time running next_pos_with_sort_not_in(), thus it runs faster on large binaries.
- (#1080) Jump table resolvers now support resolving ARM jump tables.
- (#1081, together with the PyVEX commit 61efbdcf6303a936aa3de35011d2d1e3fe5fdea5) The memory footprint of CFGFast is noticeably smaller, especially on large binaries (over 10 MB in size).
- (#1034) Concretizing a SimFile with unconstrained size can no longer run you out of memory.
- Other minor changes and bug fixes.

---

angr 7.8.6.16

- The modeling of file system is refactored.
- (#808) Add a new class Control flow blanket (CFBlanket) to support generating a linear view of a control flow graph.
- (#863) Add support to AIL, the new angr intermediate language (still pretty WIP though). Merged in several static analyses (reaching definition analysis, VEX-to-AIL translation, redundant assignment elimination, code region identification, conrol flow structuring, etc.) that support the development of decompilation in the near future.
- (#888) SimulationManager is extensively refactored and cleaned up.
- (#892) Keystone is integrated. You can assemble instructions inside angr now.
- (#897) A new class `PluginHub` is added. Plugins (analyses, engines) are refactored to be based on `PluginHub`.
- (#899) Support of bidirectional mapping between syscall numbers and syscalls.
- (#925, #941, #942) A bunch of library function prototypes (including glibc) are added to angr.
- (#953) Fix the issue where evaluating the jump target of a jump table that contains many entries (e.g., > 512) is extremely slow.
- (#964) State options are now stored in insances of SimStateOptions. `state.options` is no longer a set of strings.
- (#973) Add two new exploration techniques: Stochastic and unique.
- (#996) SimType structs are now much easier to use.
- (#998) Add a new state option `PRODUCE_ZERODIV_SUCCESSORS` to generate divide-by-zero successors.
- Speed improvements and bug fixes in CFG generation (CFGFast and CFGAccurate).

---

angr 7.8.2.21

- Refactor of how syscall handling and SimSyscallLibrary work - it is now possible to handle syscalls using multiple ABIs in the same process
- Added syscall name-number mappings from all linux ABIs, parsed from gdb
- Add `ManualMergepoint` exploration technique for when veritesting is too mysterious for your tastes
- Add `LoopSeer` exploration technique for managing loops during symbolic exploration (credit @tyb0807)
- Add `ProxyTechnique` exploration technique for easily composing simple lambda-based instrumentations (credit @danse-macabre)

---

angr 7.7.12.16

- You can now tell where the variables implicitly created by angr come from! `state.solver.BVS` now can take a `key` parameter, which describes its meaning in relation to the emulated environment. You can then use `state.solver.get_variables(...)` and `state.solver.describe_variables(...)` to map tags and ASTs to and from each other. Check out the API docs!
- The SimOS for a project is now a public property - `project.simos` instead of `project._simos`. Additionally, the SimOS code structure has been shuffled around a bit - it's now a subpackage instead of a submodule.
- The core components of Tracer and Driller have been refactored into Exploration Techniques and integrated into angr proper, so you can now follow instrution traces without installing another repostory! (credit @tyb0807)
- Archinfo now contains a `byte_width` parameter and angr supports emulation of platforms with non-octet bytes, lord help us
- Upgraded to networkx 2 (credit @tyb0807)
- Hopefully installation issues with capstone should be fixed FOREVER
- Minor fixes to gender

---

angr 7.7.9.8

Welcome to angr 7! We worked long and hard all summer to make this release the best ever. It introduces several breaking changes, so for a quick guide on the most common ways you'll need to update your scripts, take a look at the migration guide.

- SimuVEX has been removed and its components have been integrated into angr
- Path has been removed and its components have been integrated into SimState, notably the new

`history` state plugin

- PathGroup has been renamed to SimulationManager
- SimState and SimProcedure now have a reference to their parent Project, though it is verboten to use it in anything other than an append-only fashion
- A new class SimLibrary is used to track SimProcedure and metadata corresponding to an individual shared library
- Several CLE interfaces have been refactored up for consistency
- Hook has been removed. Hooking is now done with individual SimProcedure instances, which are shallow-copied at execution time for thread-safety.
- The `state.solver` interface has been cleaned up drastically

These are the major refactor-y points. As for the improvements:

- Greatly improved support for analyzing 32 bit windows binaries (partial credit @schieb)
- Unicorn will now stop for stop points and breakpoints in the middle of blocks (credit @bennofs)
- The processor flags for a state can now be accessed through `state.regs.eflags` on x86 and `state.regs.flags` on ARM (partial credit @tyb0807)
- Fledgling support for emulating exception handling. Currently the only implementation of this is support for Structured Exception Handling on Windows, see `angr.SimOS.handle_exception` for details
- Fledgling support for runtime library loading by treating the CLE loader as an append-only interface, though only implemented for windows. See `cle.Loader.dynamic_load` and `angr.procedures.win32.dynamic_loading` for details.
- The knowledge base has been refactored into a series of plugins similar to SimState (credit @danse-macabre)
- The testcase-based function identifier we wrote for CGC has been integrated into angr as the Identifier analysis
- Improved support for writing custom VEX lifters

---

angr 6.7.6.9

- angr: A static data-flow analysis framework has been introduced, and implemented as part of the `ForwardAnalysis` class. Additionally, a few exemplary data-flow analyses, like `VariableRecovery` and `VariableRecoveryFast`, have been implemented in angr.
- angr: We introduced the notion of *variable* to the angr world. Now a VariableManager is available in the knowledge base. Variable information can be recovered by running a variable recovery analysis. Currently the variable information recovered for each function is still pretty coarse. More updates to it will arrive soon.
- angr: Fix a bug in the topological sorting in `CFGUtils`, which resulted in suboptimal graph node ordering after sorting.
-

SimuVEX: `LAZY_SOLVES` is no longer enabled by default during symbolic execution. It's still there if it's wanted, but it just caused confusion when on by default.

- SimuVEX: Thanks to @ekilmer, a few new libc SimProcedures are added.
- SimuVEX: The default memory model has been refactored for expandability. Custom pages can now be created (derive the simuvex.storage.ListPage class) and used instead of the default page classes to implement custom memory behavior for specific pages. The user-friendly API for this is pending the next release.
- angr-management: Implemented our own graph layout and edge routing algorithm. We do not rely on grandalf anymore.
- angr-management: Added support for displaying variable information for operands.
- angr-management: Added support for highlighting dependent operands when an operand is highlighted.

---

angr 6.7.3.26

Building off of the engine changes from the last release, we have begun to extend angr to other architectures. AVR and MSP430 are in progress. In the meantime, subwire has created a reference implementation of BrainFuck support in angr, done two different ways! Check out angr-platforms for more info!

- We have rebased our fork of VEX on the latest master branch from Valgrind (as of 2 months ago, at least...). We have also submitted our patches to VEX to upstream, so we should be able to stop maintaining a fork pretty soon.
- The way we interact with VEX has changed substancially, and should speed things up a bit.
- Loading sets of binaries with many import symbols has been sped up
- Many, many improvements to angr-management, including the switch away from enaml to using pyside directly.

---

angr 6.7.1.13

For the last month, we have been working on a major refactor of the angr to change the way that angr reasons about the code that it analyzes. Until now, angr has been bound to the VEX intermediate representation to lift native code, supporting a wide range of architectures but not being very expandable past them. This release represents the ground work for what we call translation and execution engines. These engines are independent backends, pluggable into the angr framework, that will allow angr to reason about a wide range of targets. For now, we have restructured the existing VEX and Unicorn Engine support into this engine paradigm, but as we discuss in our blog post, the plan is to create engines to enable angr's reasoning of Java bytecode and source code, and to augment angr's environment support through the use of external dynamic sandboxes.

For now, these changes are mostly internal. We have attempted to maintain compatibility for end-users, but those building systems atop angr will have to adapt to the modern codebase. The following are the major changes:

- simuvex: we have introduced SimEngine. SimEngine is a base class for abstractions over native code. For example, angr's VEX-specific functionality is now concentrated in SimEngineVEX, and new engines (such as SimEngineLLVM) can be implemented (even outside of simuvex itself) to support the analysis of new types of code.

- simuvex: as part of the engines refactor, the SimRun class has been eliminated. Instead of different subclasses of SimRun that would be instantiated from an input state, engines each have a `process` function that, from an input state, produces a SimSuccessors instance containing lists of different successor states (normal, unsat, unconstrained, etc) and any engine-specific artifacts (such as the VEX statements. Take a look at `successors.artifacts`).
- simuvex: `state.mem[x:] = y` now *requires* a type for storage (for example `state.mem[x:].dword = y`).
- simuvex: the way of calling inline SimProcedures has been changed. Now you have to create a SimProcedure, and then call `execute()` on it and pass in a program state as well as the arguments.
- simuvex: accessing registers through `SimRegNameView` (like `state.regs.eax`) always triggers SimInspect breakpoints and creates new actions. Now you can access a register by prefixing its name with an underscore (e.g. `state.regs._eax` or `state._ip`) to avoid triggering breakpoints or creating actions.
- angr: the way hooks work has slightly changed, though is backwards-compatible. The new angr.Hook class acts as a wrapper for hooks (SimProcedures and functions), keeping things cleaner in the `project._sim_procedures` dict.
- angr: we have deprecated the keyword argument `max_size` and changed it to to `size` in the `angr.Block` constructor (i.e., the argument to `project.factory.block` and more upstream methods (`path.step`, `path_group.step`, etc).
- angr: we have deprecated `project.factory.sim_run` and changed it to to `project.factory.successors`, and it now generates a `SimSuccessors` object.
- angr: `project.factory.sim_block` has been deprecated and replaced with `project.factory.successors(default_engine=True)`.
- angr: angr syscalls are no longer hooks. Instead, the syscall table is now in `project._simos.syscall_table`. This will be made "public" after a usability refactor. If you were using `project.is_hooked(addr)` to see if an address has a related SimProcedure, now you probably want to check if there is a related syscall as well (using `project._simos.syscall_table.get_by_addr(addr) is not None`).
- pyvex: to support custom lifters to VEX, pyvex has introduced the concept of backend lifters. Lifters can be written in pure python to produce VEX IR, allowing for extendability of angr's VEX-based analyses to other hardware architectures.

As usual, there are many other improvements and minor bugfixes.

- claripy: support `unsat_core()` to get the core of unsatness of constraints. It is in fact a thin wrapper of the `unsat_core()` function provided by Z3. Also a new state option `CONSTRAINT_TRACKING_IN_SOLVER` is added to SimuVEX. That state option must be enabled if you want to use `unsat_core()` on any state.

- simuvex: `SimMemory.load()` and `SimMemory.store()` now takes a new parameter `disable_actions`. Setting it to True will prevent any SimAction creation.
- angr: CFGFast has a better support for ARM binaries, especially for code in THUMB mode.
- angr: thanks to an improvement in SimuVEX, CFGAccurate now uses slightly less memory than before.
- angr: `len()` on path `trace` or `addr_trace` is made much faster.
- angr: Fix a crash during CFG generation or symbolic execution on platforms/architectures with no syscall defined.
- angr: as part of the refactor, `BackwardSlicing` is temporarily disabled. It will be re-enabled once all DDG-related refactor are merged to master.

Additionally, packaging and build-system improvements coordinated between the angr and Unicorn Engine projects have allowed angr's Unicorn support to be built on Windows. Because of this, `unicorn` is now a dependency for `simuvex`.

Looking forward, angr is poised to become a program analysis engine for binaries *and more*!

---

angr 5.6.12.3

It has been over a month since the last release 5.6.10.12. Again, we’ve made some significant changes and improvements on the code base.

- angr: Labels are now stored in KnowledgeBase.
- angr: Add a new analysis: `Disassembly`. The new Disassembly analysis provides an easy-to-use interface to render assembly of functions.
- angr: Fix the issue that `ForwardAnalysis` may prematurely terminate while there are still un-processed jobs.
- angr: Many small improvements and bug fixes on `CFGFast`.
- angr: Many small improvements and bug fixes on `VFG`. Bring back widening support. Fix the issue that `VFG` may not terminate under certain cases. Implement a new graph traversal algorithm to have an optimal traversal order. Allow state merging at non-merge-points, which allows faster convergence.
- angr-management: Display a progress during initial CFG recovery.
- angr-management: Display a “Load binary” window upon binary loading. Some analysis options can be adjusted there.
- angr-management: Disassembly view: Edge routing on the graph is improved.
- angr-management: Disassembly view: Support starting a new symbolic execution task from an arbitrary address in the program.
- angr-management: Disassembly view: Support renaming of function names and labels.
- angr-management: Disassembly view: Support “Jump to address”.
- angr-management: Disassembly view: Display resolved and unresolved jump targets. All jump targets are double-clickable.
- SimuVEX: Move region mapping from `SimAbstractMemory` to `SimMemory`. This will allow an

easier conversion between `SimAbstractMemory` and `SimSymbolicMemory`, which is to say, conversion between symbolic states and static states is now possible.

- SimuVEX & claripy: Provide support for `unsat_core` in Z3. It returns a set of constraints that led to unsatness of the constraint set on the current state.
- archinfo: Add a new Boolean variable `branch_delay_slot` for each architecture. It is set to True on MIPS32.

---

angr 5.6.8.22

Major point release! An incredible number of things have changed in the month run-up to the Cyber Grand Challenge.

- Integration with Unicorn Engine supported for concrete execution. A new SimRun type, SimUnicorn, may step through many basic blocks at once, so long as there is no operation on symbolic data. Please use our fork of unicorn engine, which has many patches applied. All these patches are pending merge into upstream.
- Lots of improvements and bug fixes to CFGFast. Rumors are angr’s CFG was only "optimized" for x86-64 binaries (which is really because most of our test cases are compiled as 64-bit ELFs). Now it is also “optimized” for x86 binaries :) (editor's note: angr is built with cross-architecture analysis in mind. CFG construction is pretty much the only component which has architecture-specific behavior.)
- Lots of improvements to the VFG analysis, including speed and accuracy. However, there is still a lot to be done.
- Lots of speed optimizations in general - CFGFast should be 3-6x faster under CPython with much less memory usage.
- Now data dependence graph gives you a real dependence graph between variable definitions. Try `data_graph` and `simplified_data_graph` on a DDG object!
- New state option `simuvex.o.STRICT_PAGE_ACCESS` will cause a `SimSegfaultError` to be raised whenever the guest reads/writes/executes memory that is either unmapped or doesn't have the appropriate permissions.
- Merging of paths (as opposed to states) is performed in a much smarter way.
- The behavior of the `support_selfmodifying_code` project option is changed: Before, this would allow the state to be used as a fallback source of instruction bytes when no backer from CLE is available. Now, this option makes instruction lifting use the state as the source of bytes always. When the option is disabled and execution jumps outside the normal binary, the state will be used automatically.
- *Actually* support self-modifying code - if a basic block of code modifies itself, the block will be re-lifted before the next instruction starts.
- Syscalls are handled differently now - Before you would see a SimRun for a syscall helper, now you'll just see a SimProcedure for the given syscall. Additionally, each syscall has its own address in a "syscalls segment", and syscalls are treated as jumps to this segment. This simplifies a lot of things analysis-wise.

- CFGAccurate accepts a `base_graph` keyword to its constructor, e.g. `CFGFast().graph`, or even `.graph` of a function, to use as a base for analysis.

- New fast memory model for cases where symbolic-addressed reads and writes are unlikely.
- Conflicts between the `find` and `avoid` parameters to the Explorer otiegnqwvk are resolved correctly. (credit clslgrnc)
- New analysis `StaticHooker` which hooks library functions in unstripped statically linked binaries.
- `Lifter` can be used without creating an angr Project. You must manually specify the architecture and bytestring in calls to `.lift()` and `.fresh_block()`. If you like, you can also specify the architecture as a parameter to the constructor and omit it from the lifting calls.
- Add two new analyses developed for the CGC (mostly as examples of doing static analysis with angr): Reassembler and BinaryOptimizer.

---

angr 4.6.6.28

In general, there have been enormous amounts of speed improvements in this release. Depending on the workload, angr should run about twice as fast. Aside from this, there have also been many submodule-specific changes:

angr

Quite a few changes and improvements are made to `CFGFast` and `CFGAccurate` in order to have better and faster CFG recovery. The two biggest changes in `CFGFast` are jump table resolution and data references collection, respectively. Now `CFGFast` resolves indirect jumps by default. You may get a list of indirect jumps recovered in `CFGFast` by accessing the `indirect_jumps` attribute. For many cases, it resolves the jump table accurately. Data references collection is still in alpha mode. To test data references collection, just pass `collect_data_references=True` when creating a fast CFG, and access the `memory_data` attribute after the CFG is constructed.

CFG recovery on ARM binaries is also improved.

A new paradigm called an "otiegnqwvk", or an "exploration technique", allows the packaging of special logic related to path group stepping.

SimuVEX

Reads/writes to the x87 fpu registers now work correctly - there is special logic that rotates a pointer into part of the register file to simulate the x87 stack.

With the recent changes to Claripy, we have configured SimuVEX to use the composite solver by default. This should be transparent, but should be considered if strange issues (or differences in behavior) arise during symbolic execution.

Claripy

Fixed a bug in claripy where `__div__` was not always doing unsigned division, and added new methods `SDiv` and `SMod` for signed division and signed remainder, respectively.

Claripy frontends have been completely rewritten into a mixin-centric solver design. Basic frontend functionality (i.e., calling into the solver or dealing with backends) is handled by frontends (in `claripy.frontends`), and additional functionality (such as caching, deciding when to simplify, etc) is handled by frontend mixins (in `claripy.frontend_mixins`). This makes it considerably easier to customize solvers to your specific needE. For examples, look at `claripy/solver.py`.

Alongside the solver rewrite, the composite solver (which splits constraints into independent constraint sets for faster solving) has been immensely improved and is now functional and fast.

angr 4.6.6.4

Syscalls are no longer handled by `simuvex.procedures.syscalls.handler`. Instead, syscalls are now handled by `angr.SimOS.handle_syscall()`. Previously, the address of a syscall SimProcedure is the address right after the syscall instruction (e.g. `int 80h`), which collides with the real basic block starting at that address, and is very confusing. Now each syscall SimProcedure has its own address, just as a normal SimProcedure. To support this, there is another region mapped for the syscall addresses, `Project._syscall_obj`.

Some refactoring and bug fixes in `CFGFast`.

Claripy has been given the ability to handle *annotations* on ASTs. An annotation can be used to customize the behavior of some backends without impacting others. For more information, check the docstrings of `claripy.Annotation` and `claripy.Backend.apply_annotation`.

angr 4.6.5.25

New state constructor - `call_state`. Comes with a refactor to `SimCC`, a refactor to `callable`, and the removal of `PathGroup.call`. All these changes are thoroughly documented, in `angr-doc/docs/structured_data.md`

Refactor of `SimType` to make it easier to use types - they can be instanciated without a SimState and one can be added later. Comes with some usability improvements to SimMemView. Also, there's a better wrapper around PyCParser for generating SimType instances from c declarations and definitions. Again, thoroughly documented, still in the structured data doc.

`CFG` is now an alias to `CFGFast` instead of `CFGAccurate`. In general, `CFGFast` should work under most cases, and it's way faster than `CFGAccurate`. We believe such a change is necessary, and will make angr more approachable to new users. You will have to change your code from `CFG` to `CFGAccurate` if you are relying on specific functionalities that only exist in `CFGAccurate`, for example, context-sensitivity and state-preserving. An exception will be raised by angr if any parameter passed to `CFG` is only supported by `CFGAccurate`. For more detailed explanation, please take a look at the documentation of `angr.analyses.CFG`.

angr 4.6.3.28

PyVEX has a structural overhaul. The `IRExpr`, `IRStmt`, and `IRConst` modules no longer exist as submodules, and those module names are deprecated. Use `pyvex.expr`, `pyvex.stmt`, and `pyvex.const` if you need to access the members of those modules.

The names of the first three parameters to `pyvex.IRSB` (the required ones) have been changed. If you were passing the positional args to IRSB as keyword args, consider switching to positional args. The order is `data`, `mem_addr`, `arch`.

The optional parameter `sargc` to the `entry_state` and `full_init_state` constructors has been removed and replaced with an `argc` parameter. `sargc` predates being able to have claripy ASTs independent from a solver. The new system is to pass in the exact value, ast or integer, that you'd like to have as the guest program's arg count.

CLE and angr can now accept file-like streams, that is, objects that support `stream.read()` and `stream.seek()` can be passed in wherever a filepath is expected.

Documentation is much more complete, especially for PyVEX and angr's symbolic execution control components.

angr 4.6.3.15

There have been several improvements to claripy that should be transparent to users:

- There's been a refactoring of the VSA StridedInterval classes to fix cases where operations were not sound. Precision might suffer as a result, however.
- Some general speed improvements.
- We've introduced a new backend into claripy: the ReplacementBackend. This frontend generates replacement sets from constraints added to it, and uses these replacement sets to increase the precision of VSA. Additionally, we have introduced the HybridBackend, which combines this functionality with a constraint solver, allowing for memory index resolution using VSA.

angr itself has undergone some improvements, with API changes as a result:

- We are moving toward a new way to store information that angr has recovered about a program: the knowledge base. When an analysis recovers some truth about a program (i.e., "there's a basic block at 0x400400", or "the block at 0x400400 has a jump to 0x400500"), it gets stored in a knowledge-base. Analysis that used to store data (currently, the CFG) now store them in a knowledge base and can *share* the global knowledge base of the project, now accessible via `project.kb`. Over time, this knowledge base will be expanded in the course of any analysis or symbolic execution, so angr is constantly learning more information about the program it is analyzing.
- A forward data-flow analysis framework (called ForwardAnalysis) has been introduced, and the CFG was rewritten on top of it. The framework is still in alpha stage - expect more changes to be made.

Documentation and more details will arrive shortly. The goal is to refactor other data-flow analysis, like CFGFast, VFG, DDG, etc. to use ForwardAnalysis.

- We refactored the CFG to a) improve code readability, and b) eliminate some bad designs that linger due to historical reasons.

---

angr 4.5.12.?

Claripy has a new manager for backends, allowing external backends (i.e., those implemented by other modules) to be used. The result is that `claripy.backend_concrete` is now `claripy.backends.concrete`, `claripy.backend_vsa` is now `claripy.backends.vsa`, and so on.

---

angr 4.5.12.12

Improved the ability to recover from failures in instruction decoding. You can now hook specific addresses at which VEX fails to decode with `project.hook`, even if those addresses are not the beginning of a basic block.

---

angr 4.5.11.23

This is a pretty beefy release, with over half of claripy having been rewritten and major changes to other analyses. Internally, Claripy has been unified -- the VSA mode and symbolic mode now work on the same structures instead of requiring structures to be created differently. This opens the door for awesome capabilities in the future, but could also result in unexpected behavior if we failed to account for something.

Claripy has had some major interface changes:

- claripy.BV has been renamed to claripy.BVS (bit-vector symbol). It can now create bitvectors out of strings (i.e., claripy.BVS(0x41, 8) and claripy.BVS("A") are identical).
- state.BV and state.BVV are deprecated. Please use state.se.BVS and state.se.BVV.
- BV.model is deprecated. If you're using it, you're doing something wrong, anyways. If you really need a specific model, convert it with the appropriate backend (i.e., claripy.backend_concrete.convert(bv)).

There have also been some changes to analyses:

- Interface: CFG argument `keep_input_state` has been renamed to `keep_state`. With this option enabled, both input and final states are kept.
- Interface: Two arguments `cfg_node` and `stmt_id` of `BackwardSlicing` have been deprecated. Instead, `BackwardSlicing` takes a single argument, `targets`. This means that we now support

slicing from multiple sources.
- Performance: The speed of CFG recovery has been slightly improved. There is a noticeable speed improvement on MIPS binaries.
- Several bugs have been fixed in DDG, and some sanity checks were added to make it more usable.

And some general changes to angr itself:

- StringSpec is deprecated! You can now pass claripy bitvectors directly as arguments.

Migrating to angr 9.1

angr 9.1 is here!

Calling Conventions and Prototypes

The main change motivating angr 9.1 is this large refactor of SimCC. Here are the breaking changes:

SimCCs can no longer be customized

If you were using the `sp_delta`, `args`, or `ret_val` parameters to SimCC, you should use the new class `SimCCUsercall`, which lets (requires) you to be explicit about the locations of each argument.

Passing SimTypes is now mandatory

Every method call on SimCC which interacts with typed data now requires a SimType to be passed in. Previously, the use of `is_fp` and `size` was optional, but now these parameters will no longer be accepted and a `SimType` will be required.

This has some fairly non-intuitive consequences - in order to accommodate more esoteric calling conventions (think: passing large structs by value via an "invisible reference") you have to specify a function's return type before you can extract any of its arguments.

Additionally, some non-cc interfaces, such as `call_state` and `callable` and `SimProcedure.call()`, now *require* a prototype to be passed to them. You'd be surprised how many bugs we found in our own code from enforcing this requirement!

PointerWrapper has a new parameter

Imagine you're passing something into a function which has a parameter of type `char*`. Is this a pointer to a single char or a pointer to an array of chars? The answer changes how we typecheck the values you pass in. If you're passing a PointerWrapper wrapping a large value which should be treated as an array of chars, you should construct your pointerwrapper as `PointerWrapper(foo, buffer=True)`. The buffer argument to PointerWrapper now instructs SimCC to treat the data to be serialized as an array of the child

`func_ty` -> `prototype`

Every usage of the name func_ty has been replaced with the name prototype. This was done for consistency between the static analysis code and the dynamic FFI.

Migrating to angr 8

angr has moved from python 2 to python 3! We took this opportunity of a major version bump to make a few breaking API changes that improve quality-of-life.

What do I need to know for migrating my scripts to python 3?

To begin, just the standard py3k changes, the relevant parts of which we'll rehash here as a reference guide:

- Strings and bytestrings
  - Strings are now unicode by default, a new `bytes` type holds bytestrings
  - Bytestring literals can be constructued with the b prefix, like `b'ABCD'`
  - Conversion between strings and bytestrings happens with `.encode()` and `.decode()`, which use utf-8 as a default. The `latin-1` codec will map byte values to their equivilant unicode codepoints
  - The `ord()` and `chr()` functions operate on strings, not bytestrings
  - Enumerating over or indexing into bytestrings produces an unsigned 8 bit integer, not a 1-byte bytestring
  - Bytestrings have all the string manipulation functions present on strings, including `join`, `upper`/`lower`, `translate`, etc
  - `hex` and `base64` are no longer string encoding codecs. For hex, use `bytes.fromhex()` and `bytes.hex()`. For base64 use the `base64` module.
- Builtin functions
  - `print` and `exec` are now builtin functions instead of statements
  - Many builtin functions previously returning lists now return iterators, such as `map`, `filter`, and `zip`. `reduce` is no longer a builtin; you have to import it from `functools`.
- Numbers
  - The `/` operator is explicitly floating-point division, the `//` operator is expliclty integer division. The magic functions for overriding these ops are `__truediv__` and `__floordiv__`
  - The int and long types have been merged, there is only int now
- Dictionary objects have had their `.iterkeys`, `.itervalues`, and `.iteritems` methods removed, and then non-iter versions have been made to return efficient iterators
- Comparisons between objects of very different types (such as between strings and ints) will raise an

exception

In terms of how this has affected angr, any string that represents data from the emulated program will be a bytestring. This means that where you previously said `state.solver.eval(x, cast_to=str)` you should now say `cast_to=bytes`. When creating concrete bitvectors from strings (including implicitly by just making a comparison against a string) these should be bytestrings. If they are not they will be utf-8 converted and a warning will be printed. Symbol names should be unicode strings.

For division, however, ASTs are strongly typed so they will treat both division operators as the kind of division that makes sense for their type.

Clemory API changes

The memory object in CLE (project.loader.memory, not state.memory) has had a few breaking API changes since the bytes type is much nicer to work with than the py2 string for this specific case, and the old API was an inconsistent mess.

Before	After
`memory.read_bytes(addr, n) -> list[str]`	`memory.load(addr, n) -> bytes`
`memory.write_bytes(addr, list[str])`	`memory.store(addr, bytes)`
`memory.get_byte(addr) -> str`	`memory[addr] -> int`
`memory.read_addr_at(addr) -> int`	`memory.unpack_word(addr) -> int`
`memory.write_addr_at(addr, value) -> int`	`memory.pack_word(addr, value)`
`memory.stride_repr -> list[(start, end, str)]`	`memory.backers() -> iter[(start, bytearray)]`

Additionally, `pack_word` and `unpack_word` now take optional `size`, `endness`, and `signed` parameters. We have also added `memory.pack(addr, fmt, *data)` and `memory.unpack(addr, fmt)`, which take format strings for use with the `struct` module.

If you were using the `cbackers` or `read_bytes_c` functions, the conversion is a little more complicated - we were able to remove the split notion of "backers" and "updates" and replaced all backers with bytearrays that we mutate, so we can work directly with the backer objects. The `backers()` function iterates through all bottom-level backer objects and their start addresses. You can provide an optional address to the function, and it will skip over all backers that end before that address.

Here is some sample code for producing a C-pointer to a given address:

```python
import cffi, cle
ffi = cffi.FFI()
ld = cle.Loader('/bin/true')


addr = ld.main_object.entry
try:
    backer_start, backer = next(ld.memory.backers(addr))
except StopIteration:
    raise Exception("not mapped")

if backer_start > addr:
    raise Exception("not mapped")

cbacker = ffi.from_buffer(backer)
addr_pointer = cbacker + (addr - backer_start)
```

You should not have to use this if you aren't passing the data to a native library - the normal load methods should now be more than fast enough for intensive use.

CLE symbols changes

Previously, your mechanisms for looking up symbols by their address were `loader.find_symbol()` and `object.symbols_by_addr`, where there was clearly some overlap. However, `symbols_by_addr` stayed because it was the only way to enumerate symbols in an object. This has changed! `symbols_by_addr` is deprecated and here is now `object.symbols`, a sorted list of Symbol objects, to enumerate symbols in a binary.

Additionally, you can now enumerate all symbols in the entire project with `loader.symbols`. This change has also enabled us to add a `fuzzy` parameter to `find_symbol` (returns the first symbol before the given address) and make the output of `loader.describe_addr` much nicer (shows offset from closest symbol).

Deprecations and name changes

- All parameters in cle that started with `custom_` - so, `custom_base_addr`, `custom_entry_point`, `custom_offset`, `custom_arch`, and `custom_ld_path` - have had the `custom_` removed from the beginning of their names.
- All the functions that were deprecated more than a year ago (at or before the angr 7 release) have been removed.
- `state.se` has been deprecated. You should have been using `state.solver` for the past few years.
- Support for immutable simulation managers has been removed. So far as we're aware, nobody was actually using this, and it was making debugging a pain.

Migrating to angr 7

The release of angr 7 introduces several departures from long-standing angr-isms. While the community has created a compatibility layer to give external code written for angr 6 a good chance of working on angr 7, the best thing to do is to port it to the new version. This document serves as a guide for this.

SimuVEX is gone

angr versions up through angr 6 split the program analysis into two modules: `simuvex`, which was responsible for analyzing the effects of a single piece of code (whether a basic block or a SimProcedure) on a program state, and `angr`, which aggregated analyses of these basic blocks into program-level analysis such as control-flow recovery, symbolic execution, and so forth. In theory, this would encourage for the encapsulation of block-level analyses, and allow other program analysis frameworks to build upon `simuvex` for their needs. In practice, no one (to our knowledge) used `simuvex` without `angr`, and the separation introduced frustrating limitations (such as not being able to reference the history of a state from a SimInspect breakpoint) and duplication of code (such as the need to synchronize data from `state.scratch` into `path.history`).

Realizing that SimuVEX wasn't a usable independent package, we brainstormed about merging it into angr and further noticed that this would allow us to address the frustrations resulting from their separation.

All of the SimuVEX concepts (SimStates, SimProcedures, calling conventions, types, etc) have been migrated into angr. The migration guide for common classes is bellow:

Before	After
simuvex.SimState	angr.SimState
simuvex.SimProcedure	angr.SimProcedure
simuvex.SimEngine	angr.SimEngine
simuvex.SimCC	angr.SimCC

And for common modules:

Before	After
simuvex.s_cc	angr.calling_conventions
simuvex.s_state	angr.sim_state
simuvex.s_procedure	angr.sim_procedure
simuvex.plugins	angr.state_plugins

simuvex.engines	angr.engines
simuvex.concretization_strategies	angr.concretization_strategies

Additionally, `simuvex.SimProcedures` has been renamed to `angr.SIM_PROCEDURES`, since it is a global variable and not a class. There have been some other changes to its semantics, see the section on SimProcedures for details.

Removal of angr.Path

In angr, a Path object maintained references to a SimState and its history. The fact that the history was separated from the state caused a lot of headaches when trying to analyze states inside a breakpoint, and caused overhead in synchronizing data from the state to its history.

In the new model, a state's history is maintained in a SimState plugin: `state.history`. Since the path would now simply point to the state, we got rid of it. The mapping of concepts is roughly as follows:

Before	After
path	state
path.state	state
path.history	state.history
path.callstack	state.callstack
path.trace	state.history.descriptions
path.addr_trace	state.history.bbl_addrs
path.jumpkinds	state.history.jumpkinds
path.guards	state.history.jump_guards
path.targets	state.history.jump_targets
path.actions	state.history.actions
path.events	state.history.events
path.recent_actions	state.history.recent_actions
path.reachable	state.history.reachable()

An important behavior change about `path.actions` and `path.recent_actions` - actions are no longer tracked by default. If you would like them to be tracked again, please add `angr.options.refs` to your state.

Path Group -> Simulation Manager

Since there are no paths, there cannot be a path group. Instead, we have a Simulation Manager now (we recommend using the abbreviation "simgr" in places you were previously using "pg"), which is exactly the same as a path group except it holds states instead of paths. You can make one with `project.factory.simulation_manager(...)`.

Errored Paths

Before, error resilience was handled at the path level, where stepping a path that caused an error would return a subclass of Path called ErroredPath, and these paths would be put in the `errored` stash of a path group. Now, error resilience is handled at the simulation manager level, and any state that throws an error during stepping will be wrapped in an ErrorRecord object, which is *not* a subclass of SimState, and put into the `errored` list attribute of the simulation manager, which is *not* a stash.

An ErrorRecord object has attributes for `.state` (the initial state that caused the error), `.error` (the error that was thrown), and `.traceback` (the traceback from the error). To debug these errors you can call `.debug()`.

These changes are because we were uncomfortable making a subclass of SimState, and the ErrorRecord class then has sufficiently different semantics from a normal state that it cannot be placed in a stash.

Changes to SimProcedures

The most noticeable difference from the old version to the new version is that the catalog of built-in simprocedures are no longer organized strictly according to which library they live in. Now, they are organized according to which *standards* they conform to, which helps with re-using procedures between different libraries. For instance, the old `SimProcedures['libc.so.6']` has been split up between `SIM_PROCEDURES['libc']`, `SIM_PROCEDURES['posix']`, and `SIM_PROCEDURES['glibc']`, depending on what specifications each function conforms to. This allows us to reuse the `libc` catalog in `msvcrt.dll` and the MUSL libc, for example.

In order to group SimProcedures together by libraries, we have introduced a new abstraction called the SimLibrary, the definitions for which are stored in `angr.procedures.definitions`. Each SimLibrary object stores information about a single shared library, and can contain SimProcedure implementations, calling convention information, and type information. SimLibraries are scraped from the filesystem at import time, just like SimProcedures, and placed into `angr.SIM_LIBRARIES`.

Syscalls are now categorized through a subclass of SimLibrary called SimSyscallLibrary. The API for managing syscalls through SimOS has been changed - check the API docs for the SimUserspace class.

One important implication of this change is that if you previously used a trick where you changed one of the SimProcedures present in the `SimProcedures` dict in order to change which SimProcedures would be used to hook over library functions by default, this will no longer work. Instead of `SimProcedures[lib][func_name] = proc`, you now need to say `SIM_LIBRARIES[lib].add(func_name, proc)`. But really you should just be using `hook_symbol` anyway.

Changes to hooking

The `Hook` class is gone. Instead, we now can hook with individual instances of SimProcedure objects, as opposed to just the classes. A shallow copy of the SimProcedure will be made at runtime to preserve thread safety.

So, previously, where you would have done `project.hook(addr, Hook(proc, ...))` or `project.hook(addr, proc)`, you can now do `project.hook(addr, proc(...))`. In order to use simple functions as hooks, you can either say `project.hook(addr, func)` or decorate the declaration of your function with `@project.hook(addr)`.

Having simprocedures as instances and letting them have access to the project cleans up a lot of other hacks that were present in the codebase, mostly related to the `self.call(...)` SimProcedure continuation system. It is no longer required to set `IS_FUNCTION = True` if you intend to use `self.call()` while writing a SimProcedure, and each call-return target you use will have a unique address associated with it. These addresses will be allocated lazily, which does have the side effect of making address allocation nondeterministic, sometimes based on dictionary-iteration order.

Changes to loading

The `hook_symbol` method will no longer attempt to redo relocations for the given symbol, instead just hooking directly over the address of the symbol in whatever library it comes from. This speeds up loading substancially and ensures more consistent behavior for when mixing and matching native library code and SimProcedure summaries.

The angr externs object has been moved into CLE, which will ALWAYS make sure that every dependency is resolved to something, never left unrelocated. Similarly, CLE provides the "kernel object" used to provide addresses for syscalls now.

Before	After
`project._extern_obj`	`loader.extern_object`
`project._syscall_obj`	`loader.kernel_object`

Several properties and methods have been renamed in CLE in order to maintain a more consistent and explicit API. The most common changes are listed below:

Before	After
`loader.whats_at()`	`loader.describe_addr`
`loader.addr_belongs_to_object()`	`loader.find_object_containing()`
`loader.find_symbol_name()`	`loader.find_symbol().name`

whatever the hell you were doing before to look up a symbol	`loader.find_symbol(name or addr)`
`loader.find_module_name()`	`loader.find_object_containing().provides`
`loader.find_symbol_got_entry()`	`loader.find_relevant_relocations()`
`loader.main_bin`	`loader.main_object`
`anything.get_min_addr()`	`anything.min_addr`
`symbol.addr`	`symbol.linked_addr`

Changes to the solver interface

We cleaned up the menagerie of functions present on `state.solver` (if you're still referring to it as `state.se` you should stop) and simplified it into a cleaner interface:

- `solver.eval(expression)` will give you one possible solution to the given expression.
- `solver.eval_one(expression)` will give you the solution to the given expression, or throw an error if more than one solution is possible.
- `solver.eval_upto(expression, n)` will give you up to n solutions to the given expression, returning fewer than n if fewer than n are possible.
- `solver.eval_atleast(expression, n)` will give you n solutions to the given expression, throwing an error if fewer than n are possible.
- `solver.eval_exact(expression, n)` will give you n solutions to the given expression, throwing an error if fewer or more than are possible.
- `solver.min(expression)` will give you the minimum possible solution to the given expression.
- `solver.max(expression)` will give you the maximum possible solution to the given expression.

Additionally, all of these methods can take the following keyword arguments:

- `extra_constraints` can be passed as a tuple of constraints.
  These constraints will be taken into account for this evaluation, but will not be added to the state.
- `cast_to` can be passed a data type to cast the result to.
  Currently, this can only be `str`, which will cause the method to return the byte representation of the underlying data.

  For example, `state.solver.eval(state.solver.BVV(0x41424344, 32, cast_to=str)` will return `"ABCD"`.