コードにおけるデータフローを解析し、セキュリティ問題を検出:

Facebook、Instagramを支えるPythonコードの静的解析ツール「Pysa」をOSSで公開

https://www.atmarkit.co.jp/ait/articles/2008/19/news030.html

Facebookは、Pythonコードのセキュリティやプライバシーの問題を検出するオープンソースの静的解析ツール「Pysa」の詳細を発表した。

2020年08月19日 08時00分 更新

[@IT]

Facebookは2020年8月7日（米国時間）、Pythonコードの静的解析ツール「Pysa」の詳細を発表した。

Pysaは、「Python static analyzer」を略した名称。Pythonコードのセキュリティやプライバシーの問題を検出、防止するために、オープンソースソフトウェア（OSS）として開発された。

Pysaの使いどころ

PysaはFacebookにおいて、Pythonコードが特定の社内フレームワーク（技術プライバシーポリシーに基づいて、ユーザーデータへのアクセスや、ユーザーデータの開示を防ぐ）を適切に使用しているかどうかのチェックや、Webアプリケーションの一般的なセキュリティ問題（XSSやSQLインジェクションのような）の検出などを通じて、幅広い問題検出に役立っているという。

Facebookは、同社の大規模Pythonアプリケーションのセキュリティ対策にPysaが貢献している事例の最たるものとして、同社の「Instagram」サービスを支えるサーバのコードベース開発における変更の自動解析を挙げている。

FacebookはPysaを、セキュリティ問題の検出に必要な定義の多くとともに、OSSとして公開した。同社は自社サービスでOSSのPythonサーバフレームワーク（「Django」「Tornado」など）を使っており、こうしたフレームワークを使っているプロジェクトにPysaを適用すれば、初回実行時からセキュリティ問題を検出し始めることができる。Facebookがカバーしていないフレームワークを対象にする場合は、データがサーバに入る場所をPysaに知らせる数行の構成コードを追加するだけで、Pysaを使用できるという。

なおFacebookは、Pysaを使って、OSSのPythonプロジェクトにおけるセキュリティ問題の検出、開示も行っており、実績例として「CVE-2019-19775」を挙げている。

Pysaの仕組み

Facebookは、PHPと互換性のあるOSSのプログラミング言語「Hack」のコードの静的解析ツール「Zoncolan」を開発し、Hackコードのセキュリティ問題の防止に役立ててきた。このZoncolanの成功にヒントを得て開発したというPysaは、Zoncolanと同じアルゴリズムを使って静的解析を行い、Zoncolanと一部のコードを共有している。

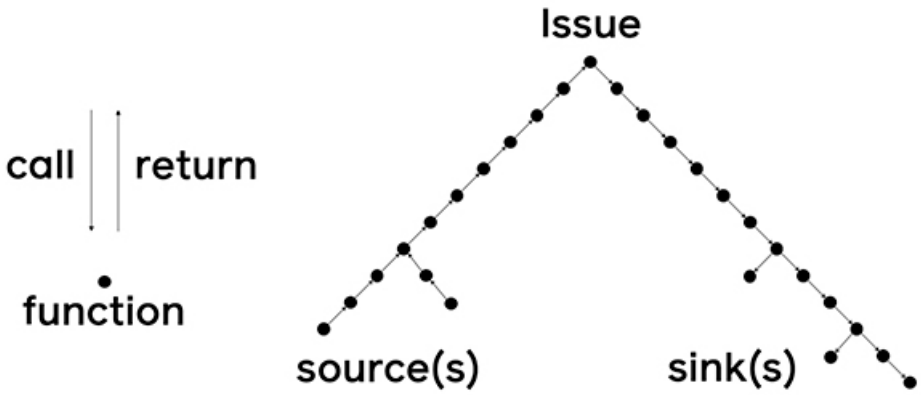
Pysaは、Facebookが開発したOSSのPython型チェッカ「Pyre」上に構築されており、Pythonコードにおけるデータフローの解析に使われている。「セキュリティ問題やプライバシー問題の多くは、不適切な場所へのデータの流入としてモデル化できるため、データフロー解析は有益だ」と同社は説明している。

Pysaを使う際にユーザーは「ソース」と「シンク」を定義する。ソースは、重要なデータの起源を指す。シンクは、ソースからのデータが到達してはならない場所だ。

セキュリティアプリケーションでは、最も一般的な種類のソースは、ユーザーが制御するデータがアプリケーションに入る場所だ。DjangoのHttpRequest.GETディクショナリがその一例だ。シンクは多くの場合、はるかに多様だが、「eval」のようなコードを実行するAPIや、「os.open」のようなファイルシステムにアクセスするAPIを含むことができる。

Pysaは反復的に解析を行い、「どの関数がソースからのデータを返すか」「どの関数が、最終的にシンクに到達するパラメーターを持つか」を識別し、要約を作成する。ソースが最終的にシンクに接続することを発見すると、問題を報告する。この

プロセスは、下図のようなツリー構造になる。



Pysa実行プロセス（出典：Facebook）

呼び出し間でのデータのフローをたどるインタープロシージャ解析には、関数呼び出しをその実装にマッピングできる必要があった。そのため、Facebookはコード内で利用可能な情報を全て利用しなければならなかった。その中には、オプションの静的型が含まれる（存在する場合）。この情報を理解するのに静的型チェッカであるPyreが使用される。PysaはPyreに大きく依存しており、両者はリポジトリを共有しているが、これらは用途が異なる別のツールだ。