PirkProposal

Pirk Proposal

Abstract

Pirk is a framework for scalable Private Information Retrieval (PIR).

Proposal

Pirk is a software framework for scalable Private Information Retrieval and is meant to provide a landing place for robust, scalable, and practical implementations of PIR algorithms. The initial scalable PIR algorithms and implementations of Pirk were developed at the National Security Agency.

Background

Private Information Retrieval (PIR) is an area of computer science and mathematics that enables a user/entity to privately and securely obtain information from a dataset, to which they have been granted access, without revealing, to the dataset owner or to an observer, any information regarding the questions asked or the results obtained. Employing homomorphic encryption techniques, PIR enables datasets to remain resident in their native locations while giving the ability to query the datasets with sensitive terms.

Rationale

Although PIR has been in existence for over twenty years, it has largely remained an academic discipline with very little robust or scalable implementation. Pirk not only provides implementations of novel scalable PIR algorithms, but it provides a framework into which robust, scalable, and practical PIR may be developed.

Pirk fits well within the Apache Software Foundation (ASF) family as it depends on numerous ASF projects and integrates with several others such as Hadoop and Spark. We also anticipate developing extensions/adaptors for several other ASF projects such as Kafka, Storm, HBase, and Accumulo in the near future.

Initial Goals

- Ensure all dependencies are compliant with Apache License version 2.0 and that all code and documentation artifacts have the correct Apache licensing markings and notice.
- Establish a formal release process and schedule, allowing for dependable release cycles in a manner consistent with the Apache development process.
- Establish a process which allows different release cycles for the core framework, extensions/adaptors, and additional algorithms.
- Grow the community to establish diversity of background and expertise.

Current Status

Meritocracy

We will actively seek help and encourage promotion of influence in the project through meritocracy. We will discuss the requirements in an open forum. We will encourage and monitor community participation so that privileges can be extended to those that contribute.

Community

Pirk currently has a community of developers within the U.S. government. In open sourcing Pirk we plan to grow the community to a broader base of industries and will work to align the interaction of our existing community.

Core Developers

The initial core developers are employed by the US Government. We will work to grow the community among a more diverse set of developers and industries.

Alignment

Pirk was developed with an open source philosophy in mind and the Apache way is consistent with the approach we have taken to date. Further, Pirk depends on numerous ASF libraries and projects including Hadoop, Spark, Commons, and Maven. We also anticipate extensions and dependencies with several more ASF projects, including Accumulo, Avro, HBase, Storm, Kafka, and others. This existing alignment with Apache and the desired community makes the Apache Incubator a good fit for Pirk.

Known Risks

Orphaned Products

Risk of orphaning is limited though it is important to grow the community. The project user and developer base is growing and there is already operational use of Pirk.

Inexperience with Open Source

The initial committers to Pirk have limited experience with true open source software development. However, despite the project origins being from closed source development we have modeled our behavior and community development on The Apache Way to the greatest extent possible. We are committed to the ideals of open source software and will eagerly seek out mentors and sponsors who can help us quickly come up to speed.

Homogenous Developers

The initial committers of Pirk come from a limited set of entities though we are committed to recruiting and developing additional committers from a broad spectrum of industries and backgrounds.

Reliance on Salaried Developers

We expect Pirk development to continue on salaried time and through volunteer time. The majority of initial committers are paid by their employers to contribute to this project. We are committed to developing and recruiting participation from developers both salaried and non-salaried.

Relationship with other Apache Projects

As described in the alignment section, Pirk is already heavily dependent on other ASF projects and we anticipate further dependence and integration with new and emerging projects in the Apache family.

An Excessive Fascination with the Apache Brand

We respect the Apache brand and desire to adopt its community building principles. Our desire is to build and foster an open source community around scalable, robust PIR which aligns with the Apache tenets. Further, Apache is a natural home for Pirk given our existing dependencies and alignment with ASF projects.

Documentation

At this time there is no Pirk documentation on the web. However, we have documentation included within the application that details usage. Using incubator infrastructure we will be rapidly expanding the available documentation to cover things like installation, developer guide, frequently asked questions, best practices, and more.

Initial Source

The core codebase is written in Java and includes detailed Javadocs and feature documentation.

Source and Intellectual Property Submission

The Pirk code and documentation materials will be submitted by the National Security Agency. Pirk has been developed by government employees. Material developed by the government employees is in the public domain and no U.S. copyright exists in works of the federal government. NSA has submitted Corporate Contributor License Agreement to the Apache Software Foundation; the Software Grant Agreement is forth coming.

External Dependencies

We believe all current dependencies are compatible with the ASF guidelines. Our dependency licenses come from the Apache v 2.0 and Eclipse Public v1.

Cryptography

Consistent with http://www.apache.org/licenses/exports/ we believe Pirk is classified as ECCN 5D002. In the event that it becomes necessary we will engage with appropriate Apache members to ensure we file any necessary paperwork or clarified any cryptographic export license concerns.

Required Resources

Mailing Lists

- dev@pirk.incubator.apache.org
- private@pirk.incubator.apache.org
- commits@pirk.incubator.apache.org

Source Control

Pirk requests use of Git for source control (git://git.apache.org/pirk.git). We request a writeable Git repo for Pirk with mirroring to be setup to Github through INFRA.

Issue Tracking

JIRA Pirk (PIRK)

Initial Committers

- Tracy Brown <tbrownpirk at gmail dot com>, CLA submitted
- Christopher Harris <Chris.Harris010 at gmail dot com>, CLA submitted
- Walter Ray-Dulaney <raydulany at gmail dot com>, CLA submitted
- Jacob Wilder <jacobwilder.opensource at gmail dot com>, CLA submitted
- Ellison Anne Williams <eawilliamsPirk at gmail dot com>, CLA confirmed
- Joe Witt <joewitt at apache dot org>, CLA confirmed

Sponsors

Champion

- Billie Rinaldi <billie at apache dot org>, IPMC Member

Nominated Mentors

- Billie Rinaldi <billie at apache dot org>, IPMC Member
- Joe Witt <joewitt at apache dot org>, IPMC Member
- Josh Elser <elserj at apache dot org>, IPMC Member
- Suneel Marthi <smarthi at apache dot org>, IPMC Member
- Tim Ellison <tellison at apache dot org>, Apache Member

Sponsoring Entity

We request the Apache Incubator to sponsor this project.