 **@ApachePirk**

Presented By: Ellison Anne Williams
Apache Pirk PPMC Member; Founder, **EN|**VEIL

ENVEIL
ENCRYPTED VEIL

Outline

What is Apache Pirk?

What is PIR?

Why Apache Pirk?

Pirk Basics

Roadmap

Get Involved

Appendix: Wideskies

What is Apache Pirk?

Framework for Scalable Private Information Retrieval (PIR)

Beautiful Blend of Mathematics & Computer Science $\mathcal{E}_g(x, y) = g^x y^N \mod N^2$

Developed at the National Security Agency

Donated to the Apache Software Foundation in July 2016

Undergoing Incubation within the Apache Incubator

Two ASF Releases To-Date – 0.3.0 Release Coming Soon

What is PIR?

PIR – Private Information Retrieval

Field of Theoretical Mathematics and Computer Science - ~20 years

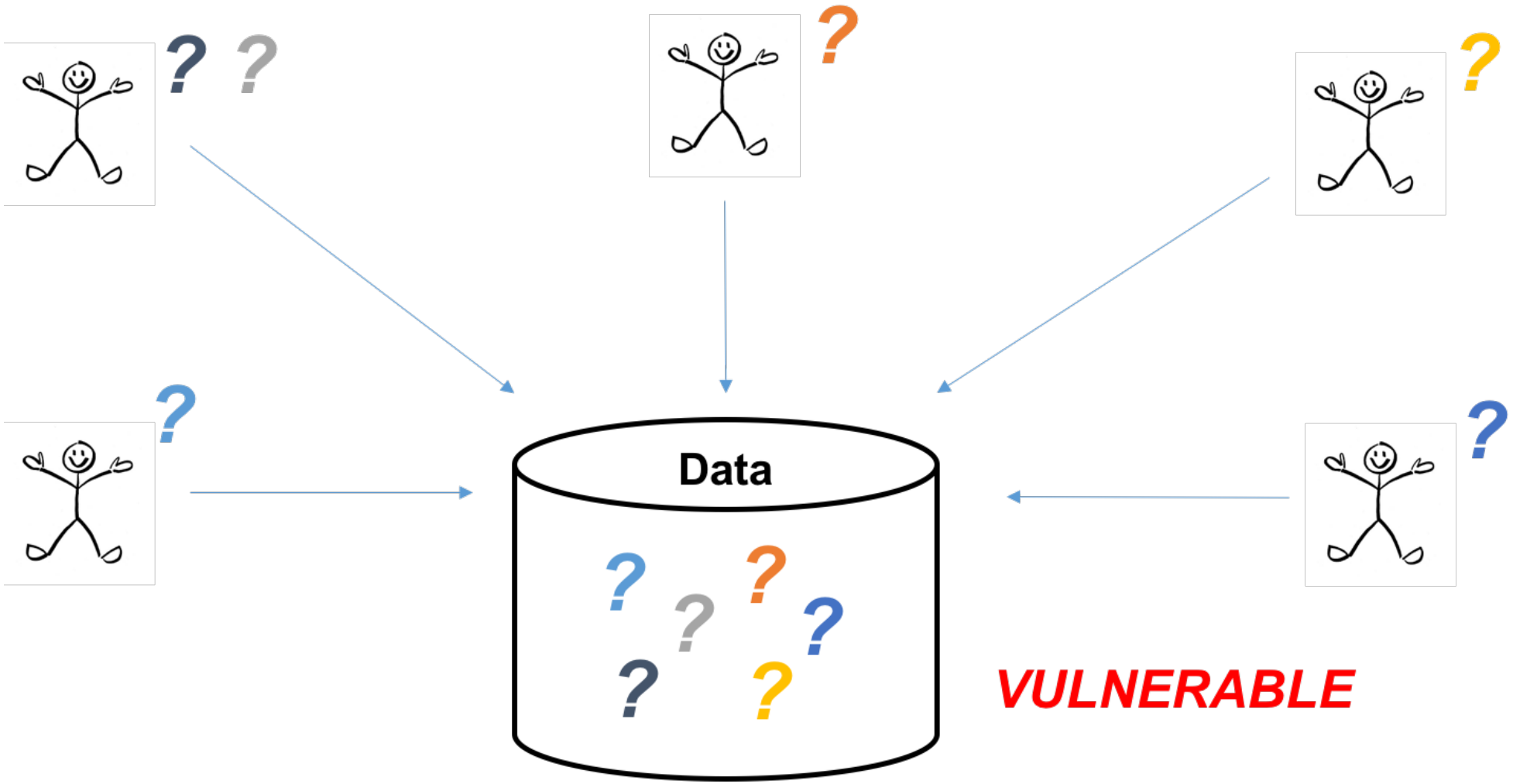
Ability to Privately Retrieve Information from a Dataset

Without Revealing Any Information Regarding the **Questions Asked** *OR* the **Results Obtained** to the Dataset Owner or an Observer

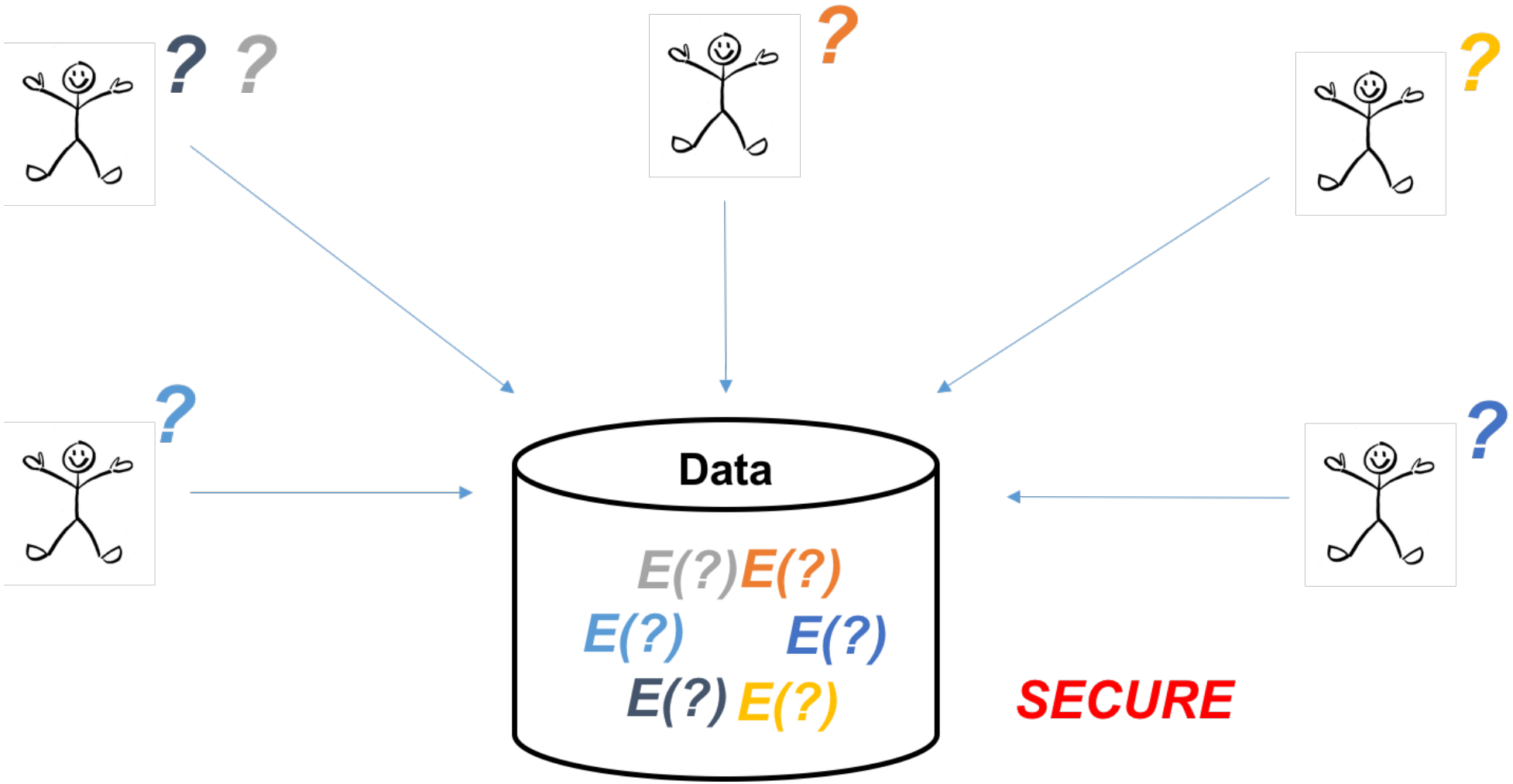
Powered by Homomorphic Encryption

$$D(\mathcal{E}(m)\mathcal{E}(m') \bmod N^2) = (m + m') \bmod N$$

Without PIR



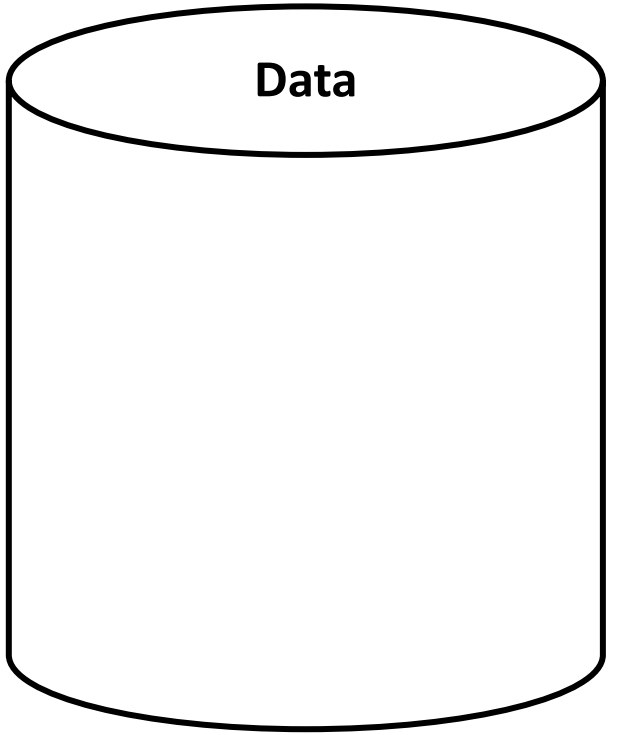
With PIR



Querier

Responder

I have a private question Q
I'm going to use PIR…

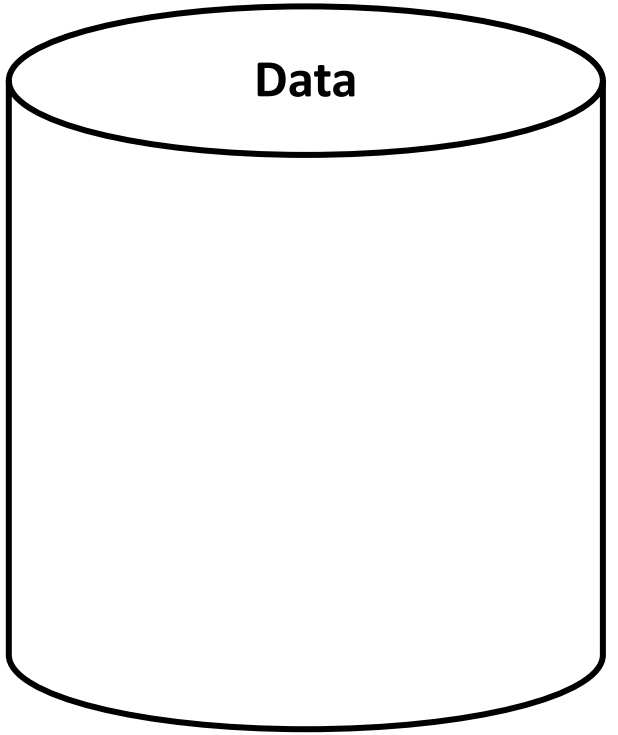


Querier

Responder

I have a private question Q
I'm going to use PIR…
I form E(Q)

E(Q)

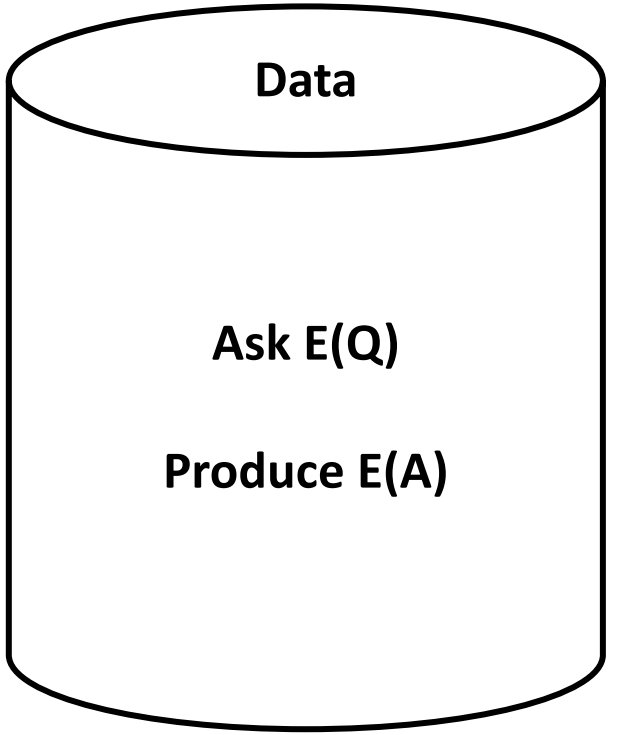


Querier

Responder

I have a private question Q
I'm going to use PIR…
I form E(Q)

E(Q)



Querier

Responder

I have a private question Q
I'm going to use PIR…
I form E(Q)

E(Q)

Data

Ask E(Q)

Produce E(A)

E(A)

Querier

Responder

I have a private question Q
I'm going to use PIR…
I form E(Q)

E(Q)

Data

Ask E(Q)

Produce E(A)

Answer A = D(E(A))

E(A)

Querier

Responder

I have a private question Q
I'm going to use PIR…
I form E(Q)

E(Q)

Data

Ask E(Q)

Produce E(A)

Answer A = D(E(A))
PIR is awesome!

E(A)

Why Apache Pirk?

PIR Historically Largely Theoretical

Need for

Practical PIR

Robust and Deployable PIR Implementations

Apache Pirk

Provides a Landing Place for Robust, Scalable PIR

Fosters a Community Around Scalable PIR

Pirk Basics

Querier

- Generates Encrypted Query Vectors
- Generates Necessary Decryption Items for Each Query Vector
- Decrypts Encrypted Results

Responder

- Performs Encrypted Queries
- Forms Encrypted Query Results

Querier

APACHE PIRK

Responder

I have a private question Q
I'm going to use PIRK…
I form E(Q)

E(Q)

Ask E(Q)

Produce E(A)

Answer A = D(E(A))
PIRK is awesome!

E(A)

Beyond the Querier and Responder

Encryption Library

- Paillier Cryptosystem Currently Implemented

Data Schema Framework

Query Schema Framework

Generic Data Filter

Testing – Distributed and In-Memory Test Suites

Data Schema

{"**date**":"2016-02-20T23:29:05.000Z",

"**src_ip**":"55.55.55.55",

"**event_type**":"dns-hostname-query",

"**query_id**":"9cef5344-3dee-41f9aa32da72d9f74778",

"**qtype**":[1,0],

"**dest_ip**":"1.2.3.6",

"**ip**":["10.20.30.40","10.20.30.60"],

"**qname**":"a.b.c.com",

”**rcode**":0}

<schema>

<schemaName> name of the schema **</schemaName>**

<element>

<name> element name **</name>**

<type> class name or type name (if Java primitive type) of the element **</type>**

<isArray> true or false -- whether or not the schema element is an array within the data **</isArray>**

<partitioner> optional - Partitioner class for the element; defaults to primitive java type partitioner **</partitioner>**

</element>

</schema>

Data Schema

{"**date**":"2016-02-20T23:29:05.000Z",

"**src_ip**":"55.55.55.55",

"**event_type**":"dns-hostname-query",

"**query_id**":"9cef5344-3dee-41f9aa32da72d9f74778",

"**qtype**":[1,0],

"**dest_ip**":"1.2.3.6",

"**ip**":["10.20.30.40","10.20.30.60"],

"**qname**":"a.b.c.com",

”**rcode**":0}

<schema>
<schemaName> awesomeDataSchema **</schemaName>**
<element>
<name> date **</name>**
<type> string **</type>**
<isArray> false **</isArray>**
<partitioner> org.apache.pirk.schema.data.partitioner.
PrimitiveTypePartitioner**</partitioner>**
</element>
…. Lots more elements ….
</schema>

Query Schema

{"**date**":"2016-02-20T23:29:05.000Z",

"**src_ip**":"55.55.55.55",

"**event_type**":"dns-hostname-query",

"**query_id**":"9cef5344-3dee-41f9aa32da72d9f74778",

"**qtype**":[1,0],

"**dest_ip**":"1.2.3.6",

"**ip**":["10.20.30.40","10.20.30.60"],

"**qname**":"a.b.c.com",

”**rcode**":0}

<schema>
<schemaName> myAwesomeQuerySchema **</schemaName>**
<dataSchemaName> superAwesomeDataSchema **</dataSchemaName>**
<selectorName> name of the element in the data schema that will be the selector **</ selectorName >**
<elements>
<name> element name **</name>**
</element>
<filterNames>
<name> (optional) element name of element in the data schema to apply pre-processing filters **</name>**
</filterNames>
<additional> (optional) additional fields for the query schema, in <key,value> pairs
<field>
<key> key corresponding the the field **</key>**
<value> value corresponding to the field **</value>**
</field>
</additional>
</schema>

Query Schema

{"**date**":"2016-02-20T23:29:05.000Z",

"**src_ip**":"55.55.55.55",

"**event_type**":"dns-hostname-query",

"**query_id**":"9cef5344-3dee-41f9aa32da72d9f74778",

"**qtype**":[1,0],

"**dest_ip**":"1.2.3.6",

"**ip**":["10.20.30.40","10.20.30.60"],

"**qname**":"a.b.c.com",

”**rcode**":0}

<schema>

<schemaName> myAwesomeQuerySchema

</schemaName>

<dataSchemaName> superAwesomeDataSchema

</dataSchemaName>

<selectorName> qname **</ selectorName >**

<elements>

<name> src_ip **</name>**

<name> dest_ip **</name>**

</element>

<filterNames>

<name> google.com **</name>**

</filterNames>

</schema>

Algorithms & Implementations

Algorithms

- Wideskies with Paillier

Querier

- Standalone, Multi-threaded

Algorithms & Implementations

- Responder
  - Standalone, Multithreaded
  - Distributed Batch
    - MapReduce, Spark
    - Data from HDFS, Elasticsearch
  - Distributed Streaming
    - Storm, Spark Streaming
    - Data from Kafka

Roadmap

Implementation Roadmap

- Input Adaptors - NoSQL Databases: Hbase, Accumulo; Kafka, Nifi
- Streaming - Storm and Heron, Spark Streaming, Flink
- Batch – Flink, Beam

Algorithmic Roadmap

- Secure Multiparty Computation, Private Set Intersection
- Fully Homomorphic Encryption

$$D(\mathcal{E}(m)\mathcal{E}(m') \bmod N^2) = (m + m') \bmod N$$
$$D(\mathcal{E}(m)^k \bmod N^2) = km \bmod N,\ k \in \mathbf{N}$$

Always on the Roadmap

- Improvements/Optimizations to Existing Code
- Benchmarking

Get Involved

We ❤️ Mathematicians *and* Computer Scientists

You don't have to code to contribute!

Apache Pirk Website

http://pirk.incubator.apache.org

Mailing Lists – Submit and Discuss Ideas/Issues

Dev: dev@pirk.incubator.apache.org

Commits: commits@pirk.incubator.apache.org

@ApachePirk

Thanks!

eawilliams@apache.org

Wideskies Appendix