Report on Workshop on Attribution at GEC 11

Matt Bishop
Dept. of Computer Science
University of California at Davis
Davis, CA 95616-8562
bishop@cs.ucdavis.edu

Introduction

This workshop, originally intended to be held at GEC 10, was moved to GEC 11 for logistical reasons at the request of the GPO.

Jeffrey Hunker and Matt Bishop led the workshop discussion, and it covered several issues about attribution, the use of attributes, policy, and experiences with GENI.

Rationale

We began by reviewing the rationale for attribution. There are issues in the management and security of networks in general, and GENI in particular, and yet as of now no provision for adequate attribution. We do not want to have this same conversation 15–20 years in the future; we would like to deal with this problem now. Should GENI become the basis for a future network, we want to introduce subject of attribution, what attribution framework would be appropriate and most useful in the context of the GENI structure, and would best support the existing and planned initiatives on GENI.

Some data simply cannot be attributed to entities; otherwise, there would be no anonymity possible. Other data must be attributable, to ensure the smooth function of the network and to trace network problems. Note that this glosses over what constitutes a disruption. To one group, for example, certain types of content may be objectionable and therefore an attack to be attributed to the attacker. Another group might consider the content to be free speech, and anonymity to be perfectly acceptable.

The concepts of attribution are more subtle than often thought. Other projects are working with attribution, and the terminology they use varies. One of our goals is to define a terminology, and use an ontology to map terms from other projects to a common vocabulary so that projects can communicate easily, and we can work with them. This will help integrate the projects with one another, and with GENI.

Thus, the rationale for developing a framework for managing attributes and their values is not immediate, but medium- to long-term. It will provide a unified view of attributes and attribution, as well as code to manage attributes. This will include APIs so those projects choosing to use the framework will be able to interact with it in a simple and uniform manner. In addition, the unified view of attributes will provide a common vocabulary for working with attribution, and for attributes; this will require an ontology for those projects using their own vocabulary.

Attribution, Privacy, and the Attribution Framework

We define attribution to be data associated with an entity. This conforms to most common views of attribution. For example, a painting is ascribed to a particular artist, and a novel to a particular author. In some cases, attribution may refer to a location at which the art was created, or (more specifically for computer security) the location and identity of an attacker.

Privacy is a key aspect of attribution: with attribution of identity, or quasi-identifiers, the privacy of the entity is compromised. Thus, circumstances of attribution, and the consequences of attributing data and values to a particular entity or set of entities, must be considered. As a result, different networks may require different levels of attribution, based upon their goals or the goals of the entities on the networks.

The general attribution framework was presented and reviewed. (See Appendix.) The key features are the inclusion of multiple entities in the interaction, including intermediate entities; the generality of what can be attributed (specifically, arbitrary data); the inclusion of a "level of assurance" (LOA) of the attribute value; and a policy negotiation system to enable entities to negotiate the types and values of attributes that they will accept.

Several different types of attribution exist: perfect non-attribution, prefect attribution, perfect selective attribution, sender non-attribution, recipient non-attribution, false attribution, randomized false attribution, imprecise attribution, and unconcern.

The issue of privacy again comes into play. Sometimes attribution is undesirable to a broad audience, but needed for a specific audience; also, sometimes it is undesirable until a specific event or set of events occurs, such as a court order enabling law enforcement to obtain the information. In that case, the attribution must be available to authorized entities, under authorized procedures, but unavailable to anyone else, and to the authorized entities before the authorized procedures are satisfied. How to do this is unclear; perhaps some sort of trusted third party or central authority would work, or some use of cryptography—but that raises issues of human as well as technological management such as key management and the use of key escrow. As technology evolves, this problem may become more tractable.

A discussion of the level of assurance of attribution followed. The LOA is not fixed and unchanging, nor will everyone necessarily agree to a value for the LOA for an attribute bound to an entity. That is, an attribute of an entity may be credible to one entity, but not credible to another. Indeed, the trust placed in the entity, and in the assigner of the value of the attribute, plays critically into the LOA—and different other entities may view those differently. Of course, there will be multiple levels of assurance; indeed, the result may be an amorphous structure rather than a hierarchy of levels.

Another factor affecting the implementation and, indeed design, of an attribution framework lies in the GENI environment. A programmer may be able to control attribution on the control plane but not on the data plane. The policy specifications need to take this into account. Indeed, a transaction may lie in both planes because it affects how data is moved or managed (control plane) and the actual data movement (data plane). We have to think in terms of how policies are specified and determined, whether based on a stated or an unstated (explicit or implicit) policy, or a negotiated policy.

Policies

One of the key benefits of an attribution framework is to make policies as explicit as possible, so entities that might be involved in a communication can determine whether they are able to participate. This raises several issues, ranging from legal to policy composition.

Legal jurisdictions map to this scheme because they govern the locations transited by the messages, as well as the locations of the sender, recipient, and intermediate entities. So, for example, a message with content objectionable to the laws of a jurisdiction in which an intermediate node resides would not be able to transit that jurisdiction, and hence that node—so either the message must use a different path, or it may not be sent. Also, different social regimes may have different attribution requirements. This might affect how a sender could broadcast (send) a message, for instance.

The above description is general. For electronic mail, attribution of sender is straightforward. But in cases (such as network-oriented programs) when individual parts cannot be attributed appropriately until all parts of the message are reassembled, the policies may prevent such messages from being sent. Indeed, entities may not be able to communicate because of conflicting attribution requirements. This would lead to a set of virtual networks over which entities could communicate, because the policies are compatible, and among which they could not communicate, because the policies of the networks are incompatible.

As an example, the problem of competing policies was raised. The sender may be bound by corporate policies as well as personal ones, and those of the receivers and intermediaries. In particular, physical intermediaries may have restrictive policies. Starbuck's may be willing to allow any traffic to transit its network (so the sender can communicate with a recipient), but Widgets R Us may restrict transiting traffic to that which furthers development and sales of widgets (in which case, a visiting sender might well be inhibited from communicating with a recipient). Thus even though a physical connection may exist, policies may inhibit its use.

It was pointed out that in practice, policy defines the *wish*, but not necessarily what actually happens. Thus, in some sense, policies are aspirational goals. But policy is tied to underlying technology and organizational capabilities, whether it is seen as "aspirational" or "feasible", and it must be crafted in light of those capabilities. In reality, just as policies constrain the technologies chosen to implement those policies, technologies also constrain (feasible, realistic) policies. This really creates a feedback loop. Note also that policy itself has layers of abstraction like operating systems and the network stack, and the difference between the oracle layer (aspirational) and feasible layer (realistic) captures the difference discussed here. Also, policy is dynamic, and this complicates the issues of working with policies.

This raises the question of the sources of the policies. Clearly, GENI must comply with the law—those policies cannot be changed. It must also comply with the policies of organizations that comprise GENI. These may be more malleable, depending on the organization's desire for involvement (and GENI's desire to have the organization involved). In practice, most of the policies seem to be "bottom up"; they are created by the practice, rather than by some entity dictating all the details of a policy upon the practitioners. As the practitioners expand their scope of work, and try new practices, conflicts with the organizational (and GENI, and other) policies will limit the practical policy.

Finally, there was some discussion about whether policies should be generally known—in some cases, one may not want a policy known—and so some metrics about what policy terms are present would be very useful.

Lessons from Other Testbeds and Networks

PlanetLab and Emulab have experience at practical attribution. The professor is in charge of her students' work. One of the systems that the students are using launches an attack on Citibank. Who is it? What level of identification is needed—to the system, to the student, to the professor, to the university, or to some other entity? This question falls directly into the attribution framework, which—when precisely implemented—would be able to provide the answer, assuming it were given the appropriate data.

PlanetLab has network flow tools to help identify the systems involved (and the slice as well).

Attribution and GENI Issues

Slides of three projects (NetKarma, ABAC, and Shibboleth) mapped the projects into our current, high-level framework. (See attached.)

The ABAC project raised issues of attribution directly. Most people think of ABAC as an authorization framework, not an attribution framework. A key issue is the policy description language, which is critical to authorization. The language can generate proofs to show what is and is not possible, to help people reason about resource authorization. The general attitude was to address what could be addressed, and not try to to address that which cannot be addressed. What scope can relevant logic languages handle?

The NetKarma project deals with provenance of measurements and experiments. Given measurements from a slice, they want to be able to make it available assuming they can get attribution to verify the data. Their provenance tool is interesting and useful. The provenance itself is value neutral; it can be used to support claims, but the nature of those claims is completely irrelevant to the provenance used to support them.

Shibboleth deals with identity attributed. The discussion involving Shibboleth emphasized the importance of LOA, as well as how to deal with identities across federations and multiple systems in the guise of single sign on. Attributes raise similar problems.

Attribution for GENI

Jeffrey Hunker, JHA LLC

Matt Bishop, UC Davis

Carrie Gates, CA Labs

Agenda

- What we are doing
  - Generalized framework for attribution
  - Policy negotiation a key part of this
  - Benefits
- Discussion
  - Questions
  - Answers?

Caution

- Terminology varies among projects
  - So we'll define ours next

(One goal of our project is an ontology of the terminology to make life easier!)

Definition

the association of data with an entity

- This is a high-level view!
  - Approach has benefits
- Attribution (dictionary definition):
  - the ascribing of a work (as of literature or art) to a particular author or artist
  - an ascribed quality, character, or right
  - determining the identity or location of an attacker or an attacker's intermediary

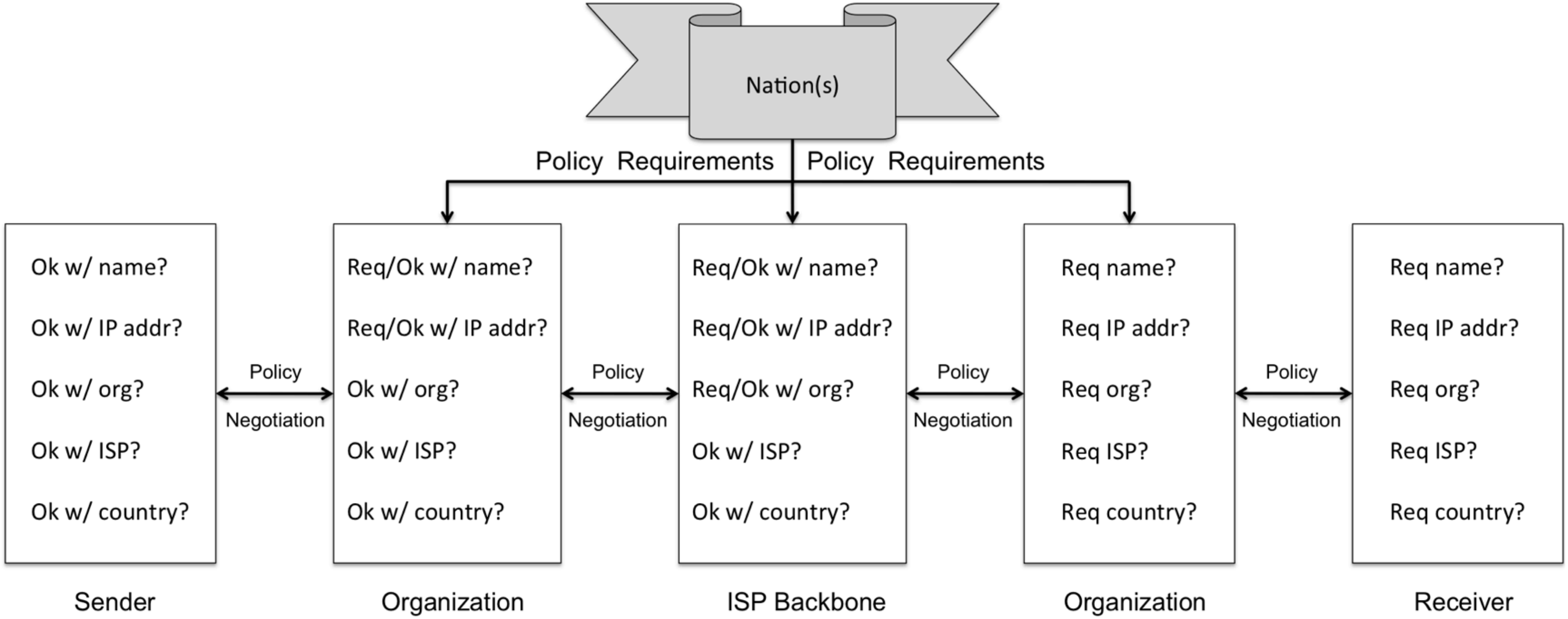
Real-Life Example: Competing/Ambiguous Needs

- "First Origin" policy
  - Technical context: net admins can track botnets to point of distribution; generally considered good
  - Political context: repressive gov'ts can track messages of dissent to point of origin; generally considered bad
- Is privacy good or bad?
  - Consider the circumstances
- Result: different networks with different levels of attribution

How We Think About It

- Level of attribution
  - Perfect non-attribution, false attribution, etc.
- Target of attribution
  - Person, IP address, organization
- Confidence in attribution
  - Attribution assurance, level of assurance (LoA)
- Adequacy of attribution
  - Depends on purpose
- Composition of attribution
  - Sender, receiver policies may vary

Attribution Framework



Sender	Organization	ISP Backbone	Organization	Receiver
Ok w/ name?	Req/Ok w/ name?	Req/Ok w/ name?	Req name?	Req name?
Ok w/ IP addr?	Req/Ok w/ IP addr?	Req/Ok w/ IP addr?	Req IP addr?	Req IP addr?
Ok w/ org?	Ok w/ org?	Req/Ok w/ org?	Req org?	Req org?
Ok w/ ISP?	Ok w/ ISP?	Ok w/ ISP?	Req ISP?	Req ISP?
Ok w/ country?	Ok w/ country?	Ok w/ country?	Req country?	Req country?

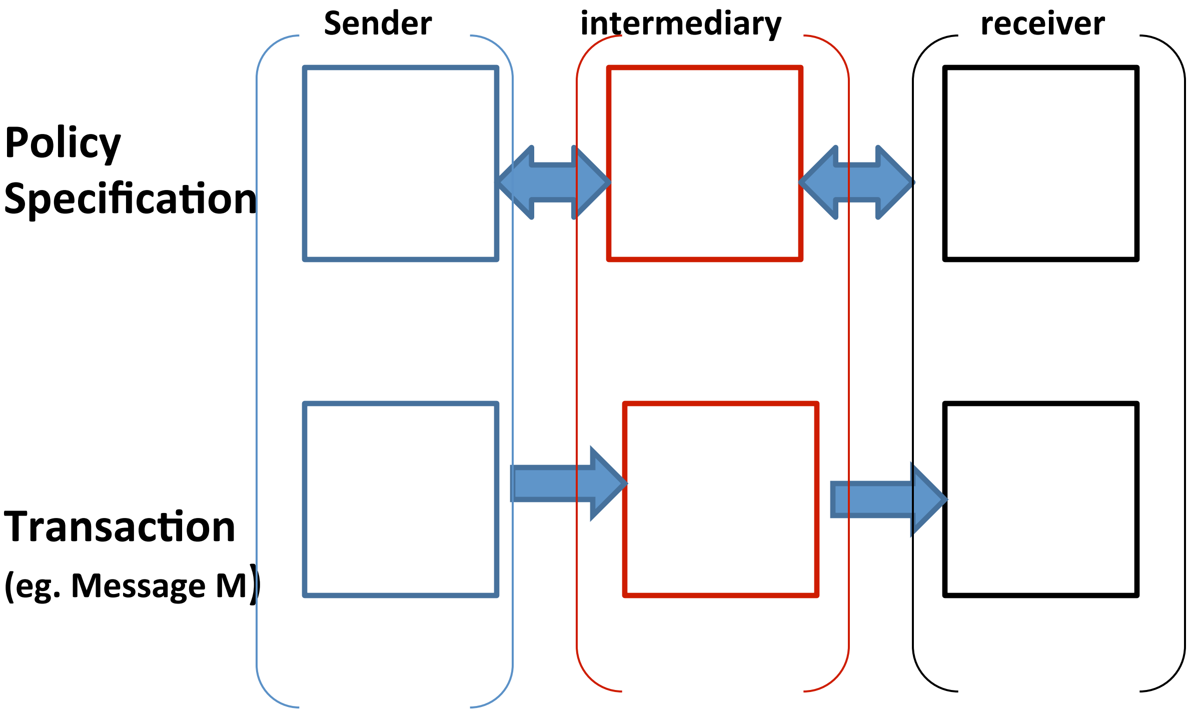
perfect non-attribution
false attribution
randomized false attribution
imprecise attribution

Set of actors
What is being attributed
Assurance of attribution
Policy negotiation system

perfect non-attribution
perfect attribution
perfect selective attribution
sender non-attribution
recipient non-attribution
unconcern

Generalized Attribution System

- Policy specification: usually *implicit*
- Transaction: what you actually do



Sender

intermediary

receiver

Policy Specification

Transaction

(eg. Message M)

Policy defines what data is tied to what entity and who has access to that data. It is determined by negotiation or agreed upon rules

Follows policy specified

Goals of Work

- Provide a unified view of attributes and attribution
  - Code to manage attributes
  - Code to help specify policy negotiation (but understanding that humans will be involved in this)
  - Ontology of terminology to help mediate and reconcile different work

Benefits

- Make assumptions explicit
  - Users of the services understand exactly what you are offering
  - You don't get criticized for not meeting what you weren't trying to do, but others thought you were
- Extensibility
  - Can adapt your services with minimal effort to work with other services and to provide higher or lower levels of authentication/identity/authorization/etc. when new folks come on line and need them
- Support your services, experiments
  - Attribution framework provides ways to negotiate policies, manage attributes
- Consistent ontology
  - So meaning of terms is clear

Other Work

- GENI projects related to attribution
  - ABAC (authorization for GENI)
  - NetKarma (provenance)
  - Shibboleth (identity management)
  - ORCA (trust structure)
  - *May be others …*

Questions

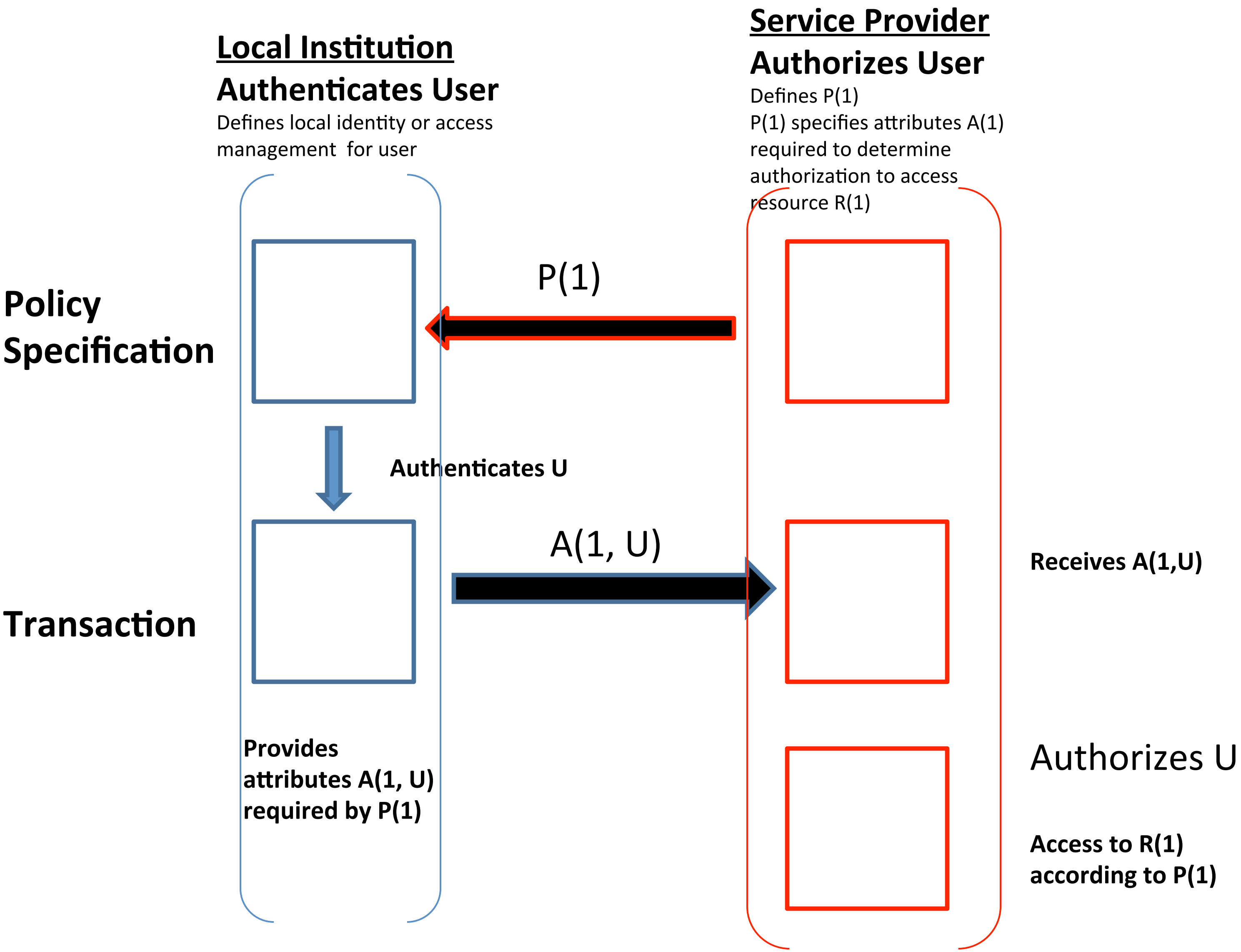
- What are the entities that you need or want attribution for?
- What sort of policies do you need for your experiments and/or services?
  - What organizational agreements are needed?
- What attributes do you need?
  - What level of assurance do you need?

Questions

- Can this view of attribution support your framework?
  - If not, what elements of an attribution framework that would help you are missing?
  - What would encourage developers to use this framework?
  - What types of attribution will be most useful to you (individual, host, organization, ISP, etc)?

Shibboleth

Authentication of User by Local Institution
Authorization for Resource Access by Service Provider



Local Institution

Authenticates User

Defines local identity or access management for user

Service Provider

Authorizes User

Defines P(1)
P(1) specifies attributes A(1) required to determine authorization to access resource R(1)

Policy Specification

P(1)

Authenticates U

A(1, U)

Receives A(1,U)

Transaction

Provides attributes A(1, U) required by P(1)

Authorizes U

Access to R(1) according to P(1)

ABAC

Attribute Based Access Control

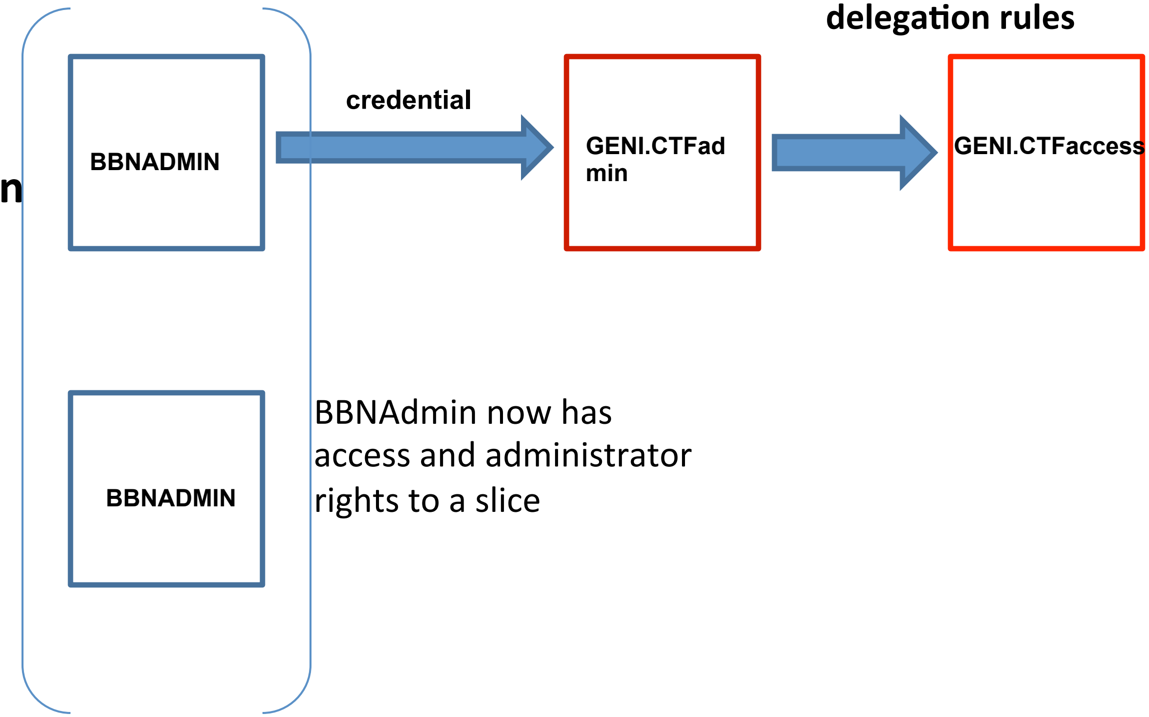
Attributes can be assigned or delegated

***Principal*: entity assigned attributes**

***Attribute*: what a principal is authorized to do (or what determines what a principal is authorized to do?)**

***Credentials:* used to assign attributes and create delegation rules**

Policy Specification



Negotiation is out of band

Transaction

BBNAdmin now has access and administrator rights to a slice

Principals not involved in transaction

NetKarma

Provenance-Based Record of Experiment

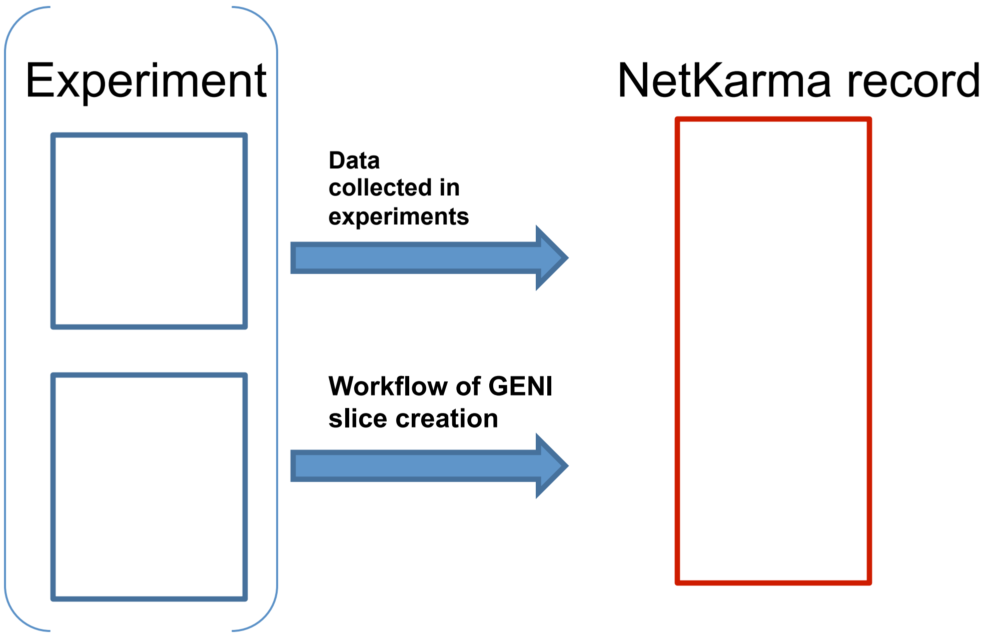
Attributes can be assigned or delegated

Policy Specification

Policy pre-specified

Negotiation is out of band

Transaction



Experiment

NetKarma record

Data collected in experiments

Workflow of GENI slice creation