2 September 2010

I. INTRODUCTION AND DEFINITION

1.1 The Problem of Attribution

Attribution has been a desired feature of networks (and, indeed, of data at rest as well, although we focus here on data in motion) for several years. Typically the approach has been a combination of IP traceback schemes in order to determine the actual IP address from which a packet was generated—a scheme that was originally designed to determine the originating IP address for spoofed packets in a denial-of-service attack—and public key infrastructure (PKI) in order to bind a particular individual to a particular message. Defenders such as security professionals and governments have traditionally defined requirements for an attribution system, and the underlying assumption is that attribution in all cases is both necessary and good. In general, attribution is desired so that someone (an individual, an organization, a nation) can be held accountable for one's actions.

However, the value of attribution must be examined from the viewpoint of multiple stakeholders. Indeed, there are cases where perfect non-attribution is desirable, such as in the case of whistleblowers or from the viewpoint of websites who will not want to provide the identities of individuals visiting their site even if compelled by subpoena. This case highlights the political and cultural aspects of attribution, because some cultures exalt the whistleblower, whereas other cultures condemn it. Further, some situations require false attribution, such as an intelligence agent being undercover and visiting a terrorist website, whereas the terrorist website might require attribution.

Further, as alluded to in the first paragraph, there are different levels of attribution, and the level required is determined by the need of the stakeholder. In some cases it might be necessary to attribute a message to a particular individual, while in other cases it might be necessary to attribute only to a specific computer, IP address, or organization. For example, arresting an individual for participating in illegal activities requires binding the individual to the activity. If a nation state has been attacked, it needs to attribute the activity to another state, and not necessarily to the specific individuals who launched the attack. Given the possible stakes inherent in the use of an attribution system, it is imperative that the system provide some indication of the degree of confidence that a user can have that the attribution is accurate and correct. Returning to the case of the individual to be arrested for illegal activity, the attribution mechanism must provide sufficient evidence and rigor to validate the attribution beyond a reasonable doubt, the standard for a criminal conviction (at least in the United States). It is not sufficient to simply provide the attribution; the attribution must be one in which the user can have confidence.

Different senders and receivers may require different attribution policies. For example, a government web site might require attribution to the user level, but be willing to negotiate down to just an IP address should the user prefer to not provide personal identity. Conversely, a dissident web site needs to advertise its policy of not accepting any forms of attribution before a visitor accidentally provides some (correct) attribution information. We therefore need to determine what policies might be

required, as well as the requirements for a negotiation system. Mechanisms for advertising policies also need to be devised, along with an examination of where policies should reside—at the end points, intermediate routers, or somewhere else.

1.2 Definition and Purpose of Attribution

The Merriam-Webster online dictionary defines "attribution" as [4]:

1. the act of attributing; especially: the ascribing of a work (as of literature or art) to a particular author or artist
2. an ascribed quality, character, or right

Specific to cyber-security needs, attribution has been defined as "determining the identity or location of an attacker or an attacker's intermediary" [5]. Within the academic literature, the term attribution (as well as accountability) tends to be used, but is not defined. In general the literature assumes the Merriam-Webster definition, with the goal of determining the person or, more commonly, the computer that originates an attack. Towards this end there has been much work in IP traceback (see, for example, Savage et al. [2] or Burch and Cheswick [1] for early work in this area) and stepping stone detection (see Staniford-Chen and Heberlein [3] for early work in this area). But we believe that this view of attribution is overly limited. The side effect of providing attribution for an attack is that attribution must also be provided for non-attack traffic[1].

We define "attribution" as the binding of data (called a *characteristic*) with an entity (person, process, file, other data, etc). For example, authentication is a mechanism for attributing an identity to an entity, and is thus an example of attribution. The time at which the data was sent is an attribute of interest in situations with temporal constraints. The route data takes over the network is an attribute that network administrators may find useful to know. It also implies how visible the data was as it goes from its source to its destination. Broadcast-style routes enable many more sites to see the data than do point-to-point routes.

The goal of attribution is to show that the characteristic associated with an entity has a particular value, or one of a particular set of values. The purpose for using attribution is generally that of accountability—in a cyber-security context attribution is generally used to determine who is initiating an attack (e.g., Wheeler and Larsen [5]) and therefore assumed to be good and desirable.

A critical observation is that the characteristic being attributed need not be identity. As noted above, in some cases the source of the data is less important than the time at which it entered a local area network. For example, the time at which an attack occurs may reveal much about the goals of the attack, even if the origin is unknown or ambiguous. Similarly, attribution is a valuable tool for network and system management for characteristics other than identity, such as times and resources used.

[1] Unless someone is able to implement the "evil bit" defined in RFC 3514 [13] successfully.

II. DEFINING THE REQUIREMENTS OF THE SENDER AND RECIPIENT

In order to define what is being attributed we need to know first who deciding what is to be attributed. Here we start with the simplest case, where the attribution is defined by the requirements of the sender and receiver. (Late we will consider the interests of others parties, and cases where attribution is provided to others than the recipient or sender).

REQUIREMENT 1: The Sender and Recipient must be defined.

In many discussions of attribution the sender of a packet is thought of as the originating machine. That may be insufficient, and in fact misleading and meaningless. As an example, consider DDoS attacks launched by 'botnets. Here, attribution back to the botnet provides little insight into the real source of the attack. Attribution may reside with the machine, the organization, and the human being. Attribution may also reside with the network. As an example, at least one major ISP (Rogers, a major ISP in Canada) has confirmed it inserts advertisements into packets responding to certain addresses [14]. These advertisements can only be attributed to the intermediate ISP, or network.

In some cases it might be necessary to attribute a message from a particular individual, while in other cases it might be necessary to attribute only from a specific computer, IP address, or organization. For example, arresting an individual for participating in illegal activities requires binding the individual to the activity. If a nation state has been attacked, it needs to attribute the activity to another state, and not necessarily to the specific individuals who launched the attack.

These same considerations apply to defining the recipients.

The specific requirements that follow from this discussion will be addressed later.

2.1 Attribution Assurance

Given the possible stakes inherent in the use of an attribution system, it is imperative that the system provide some indication of the degree of confidence that a user can have that the attribution is accurate and correct. Returning to the case of the individual to be arrested for illegal activity, the attribution mechanism must provide sufficient evidence and rigor to validate the attribution beyond a reasonable doubt, the standard for a criminal conviction (at least in the United States). It is not sufficient to simply provide the attribution; the attribution must be one in which the user can have confidence.

REQUIREMENT 2: There must be a way to define the levels of *attribute assurance* – the metrics or means of senders/recipients (or whatever nodes need this – to be called *trust client nodes*) to assess trust in the accuracy and security of the communication of the attribution characteristics.

We further define that the binding between the entity and the characteristic is '*provably attributable'* if the attribution assurance meets the standard of proof by the interested party, whatever that standard of proof might be. In a legal environment, that standard of proof would be a legal standard as defined by

the canon of legal 'proof.' In other instances, 'provably attributable' could satisfy standards required for, say, the use of force under the Law of War, or, in more benign circumstances, the standard of proof required in social relationships. Hence, the standard of 'provably attributable' is *context dependent* based on the values of the party interpreting the entity-characteristic binding.

How to structure metrics to assess trust in the attribution's accuracy and security will be discussed later in the paper (in Attribution Vector).

As already noted, a critical observation is that identity is not the only characteristic that can (or should) be attributed. However, for the following discussion, let us postpone the full discussion of the characteristic and further assume that the identity of the entity is the characteristic of attribution.

2.2 Attribution Policies of Senders and Recipients

Consider the following scenarios

Example 1: The Rock Grain Company wants to become the supplier for UC Davis' student cafeteria. The two negotiate a contract over the Internet. The final exchange involves a signed contract, sent from and signed by the UC Davis Dining Director, then received, signed, and returned by the Rock Grain Company President. Correct attribution of both signers is critical because for business purposes, both the senders and the receivers must be certain their peer is the party who may legally commit the peer's company (or institution).

Example 2: A group of attackers launch a distributed denial of service (DDoS) attack on a company that does all its business over the web. When the flooding begins, the company needs to have the flooding packets attributed to the originators of the attack. The originators of the attack, on the other hand, do not want those packets attributed to them. Here, the senders want non-attribution, but the recipients want attribution.

Example 3: Intelligence agents are examining terrorist web sites. The web sites want to know who is looking at them, both to get information about potential recruits and to know if adversaries (i.e., intelligence agents) are examining the sites for information about potential attacks. Here, the senders (the terrorist web sites) want full attribution; the recipients (the intelligence agents) want their traffic non attributable (not merely unattributed).

Example 4: Consider two dissidents in a repressive government who wish to communicate. As neither fully trusts the other, and both believe that the government may be monitoring the messages, neither wants attribution of any kind. Thus, here the sender and the receiver want no attribution.

These four scenarios present cases where attribution requirements differ.

REQUIREMENT 3: Policy Requirements of Senders and Recipients must be specified, and the attribution framework in its full form should allow for specifying a range of possible attribution policies: Analyzing this in terms of senders and recipients, the recipients may want to have any of the following forms of attribution available:

- *Perfect non-attribution*, in which attribution is not possible; for example, the dissident scenario;
- *Perfect attribution*, in which the attribution from both the sender and recipient are known to both; for example, the business scenario shown above;
- *Perfect selective attribution*, in which the recipient wants the attribution known to some entities but not to others; for example, a recipient may care that her spouse knows she received a payment, but not her employer;
- *Sender non-attribution*, in which the recipient does not want to be able to attribute characteristics to the sender; for example, a whistleblower such as "Deep Throat" in the Watergate scandal;
- *Recipient non-attribution*, in which the recipient wants to attribute characteristics to a sender but does not want the sender to be able to attribute anything to the recipient; for example, the intelligence agent scenario;
- *Unconcern*, in which the recipient does not care about attribution.

Similarly, the senders may want to disguise some of the attributes of the message in one of the following ways:

- *Perfect non-attribution*; for example, a whistleblower;
- *False attribution*, in which the recipient can perceive attribution of the message but the characteristics, while consistent, is inaccurate; for example, the intelligence agent scenario above, with the agent wanting the terrorists to attribute her messages from an ally of the terrorists;
- *Randomized false attribution*, or false attribution without the consistency; for example, the intelligence agent scenario in which the agent repeatedly visits the web site, each time under a different identity;
- *Imperfect attribution*, in which the recipient can attribute characteristics accurately, but to do so takes too long (so the knowledge is useless or redundant) or costs more than the value of knowing the attribution. Note that imperfect attribution is context dependent, based on the resources or patience of the recipient.

III. PRIVACY AND ATTRIBUTION – The Idea of 'Attribution Privacy'

As already noted, attribution is the binding of a characteristic with an entity. *The attribution privacy of entity A with respect to entity C* is when the binding of A to a characteristic is kept secret from C. Jeffrey sends a message M to Matt with perfect attribution, but the attribution that Jeffrey sent M to Matt is kept secret (in whatever way) from Carrie. Note that whether the content of M is secret or not is

immaterial to the attribution privacy of Jeffrey – what is kept secret from Carrie is the attribution of the origin of M is Jeffrey.

Using classical encryption can provide message secrecy or attribution of origin, but not both simultaneously:

Example 5: Matt and Jeffrey possess a shared secret key. Matt sends Jeffrey a message, and Jeffrey can decrypt it since both possess the secret key. But Jeffrey cannot proveto a disinterested third party (a "judge") that the attribution of the message origin is Matt. Jeffrey could have created the message himself, since Jeffrey possesses the secret key as well as Matt.

An attribution privacy violation of entity A with respect to entity C occurs when the attribution of M from A is known to C, when A desires that the attribution be kept secret from C.

Example 6: Jeffrey sends a message to Matt. Matt can attribute the message from Jeffrey. Jeffrey has attribution privacy with respect to Carrie if Matt cannot demonstrate (to Carrie) that the message is provably attributable from Jeffrey. All Carrie can accept is Matt's word that the message is attributable from Jeffrey.

Jeffrey's attribution privacy is violated if Jeffrey sends a message to Matt, but does not want anyone (including Matt) to be able to provably attribute the message to him. If Matt can provably attribute that to Carrie, then Jeffrey's attribution privacy has been violated. Whether or not Carrie knows the content of the message is immaterial to whether Jeffrey's attribution privacy has been violated.

Being able to demonstrate to others that a message is attributable from a sender may be particularly important in legal work, for example when legal services are outsourced, such as to countries like India.

Example 7: Jeffrey is a lawyer who sends a legal document to Matt. There is perfect attribution between Jeffrey and Matt, so that Matt can attribute the source of the message from Jeffrey. Matt needs also to demonstrate to Carrie (a judge) that the document is attributable from Jeffrey. In this case, the technical qualities of attribution need to be acceptable as "proof" (in this case legal proof) of the message attribution – i.e., provably attributable.

This example also demonstrates that the meaning of "provable attribution" depends upon the context in which the term is used. In a legal context (as in the example), "provable" means the standard of proof required by a court of law. In a technical context, "provable" means that the analysts are certain that the attribution is correct, but they may not be able to demonstrate it to the level of proof a court would require. For example:

Example 8: Jeffrey sends a secret document to Matt. Matt does not wish to share the contents of the document with Carrie, but needs to demonstrate to Carrie that the message containing the document is attributable from Jeffrey. In this case a signed cryptographic hash of the document would serve as provable attribution of the document's source.

In the above examples, it isn't specified if entities other than Carrie know that the message is attributed from Jeffrey. In the following examples, it does matter if entities other than Carrie can attribute the message to Jeffrey:

Example 9: Jeffrey sends a legal document to Matt and Carrie. Jeffrey wishes that Matt and Carrie attribute the message from Jeffrey, but that Matt and Carrie cannot prove this attribution (i.e., provably attribute this message) to anyone else.

Example 10: Jeffrey sends a document to Matt, but not to Carrie. Jeffrey wishes that Matt can demonstrate to Carrie that the message is attributable from Jeffrey, but to no one else.

In both of these examples, if either Matt or Carrie can demonstrate to anyone else that the message is attributable from Jeffrey, then Jeffrey's attribution privacy has been violated.

Complete anonymity of A is when A's attribution privacy extends to everyone (every entity) i.e., the binding of A to a characteristic (e.g., a message M) is kept secret from everyone.

Jeffrey has complete anonymity under the following condition: when he sends a message to Matt (or anyone else), no one can demonstrate that the message is attributable from Jeffrey.

Issue: Jeffrey may want to be able to prove that the message is attributable from himself, but without being able to prove it to anyone else, as in the following example:

Example 11: Matt sends Jeffrey a message, and Jeffrey posts it somewhere (with or without Matt's name associated with it). Matt can prove to himself that what Jeffrey posted is what Matt sent Jeffrey, but Matt cannot prove it to anyone else. Jeffrey cannot prove that the message is attributable from Matt.

3.1 Discussion about Attribution Privacy

Attribution privacy raises several issues for which the implications for attribution system requirements are uncertain.

Issue 1: Specifying the form by which the association between entity and characteristic is made

The core question seems to be "how is the association made between the entity and the characteristic (in this case, Jeffrey or Matt and the message)?" There are two alternatives (?):

The association is based on a 'thing' that can be shared and once produced is no longer in A's control– like a notary public's seal, which is under the control of the notary public and not under A's control. Jeffrey sends a message to Matt, so Matt can now share with Carrie the 'thing' that provides the attribution back to Jeffrey. Jeffrey may not want the 'thing' to be shared with Carrie (whether or not the message itself is shared), so Jeffrey's attribution privacy is violated.

The association is based on a 'quality' that cannot be shared without the consent of A. Matt has the attribution of Jeffrey to a message M, but cannot share this attribution with Carrie. Carrie has to trust that Matt is telling the truth when Matt says that M is attributed from Jeffrey. Jeffrey could lie and tell Carrie that M is not associated with Jeffrey. In this case Carrie has to decide whether to believe Jeffrey or Matt. Jeffrey's attribution privacy is not violated even by Matt's disclosure of M to Carrie, since whatever quality provides the attribution of M from Jeffrey cannot be shared.

Issue 2: Attribution privacy and perfect selective attribution

Hypothesis (Nb., the authors believe that this hypothesis could be stated and proved as a lemma with the proper mathematical framework): Attribution privacy violations only could occur when perfect selective attribution (defined above) is desired. If Jeffrey sends a message, and wants everyone to know that the message is attributed to Jeffrey, then there can be no attribution privacy violation. If Jeffrey has perfect anonymity (i.e., Jeffrey's attribution is secret from everyone) then unless there are technical violations, again no attribution privacy violation can happen. It is only when Jeffrey doesn't want that Carrie can attribute from Jeffrey the message sent to Matt (and Matt can attribute the message to Jeffrey) that there is the potential for attribution privacy violations.

Issue 3: Defining a default attribution baseline?

Is it the case that our current baseline on the Internet makes the issue of attribution privacy violations less significant That is, Jeffrey says that "Carrie can't actually perfectly attribute M to Jeffrey" but Carrie just says, "c'mon, we know that from a practical perspective most likely M came from Jeffrey"?

Stated differently, does the *default attribution baseline* of the network (if one is specified) shape the practical consequences of how important the consequences of attribution privacy and attribution privacy violations are? It is unclear exactly what the default attribution baseline for the existing Internet is, if in fact one exists, but in a future system it seems important to at least make a choice as to whether such a default baseline should exist.

IV. AN ATTRIBUTION FRAMEWORK

The discussion so far has abstracted away key aspects of the full attribution framework so as to illustrate both the policy requirements of the entity pair, and their possible need for attribution privacy.

Five aspects of attribution are relevant to our discussion of the full attribution framework:

- First is the *set of actors*.
- The second is *what is being attributed*. We represent this by an *attribute vector* that lists the characteristics for which the values are requested, or lists the pairs of characteristics and their values.

- The third aspect is *assurance*, namely confidence that those values are correct; we refer to this as the *attribute assurance*.
- The fourth aspect is exactly who is the 'sender' for the attribution, and to whom is the attribution provided? While it is often tacitly assumed that attribution from the sender is provided to the recipient, *attribution also could be provided to other entities with or without the recipient also obtaining the attribution.*
- The final component is a *policy negotiation* system that the actors use to negotiate an acceptable level of attribute assurance, or to determine that no such level is possible under the extant circumstances.

Figure 1 provides an overview of our attribution framework.

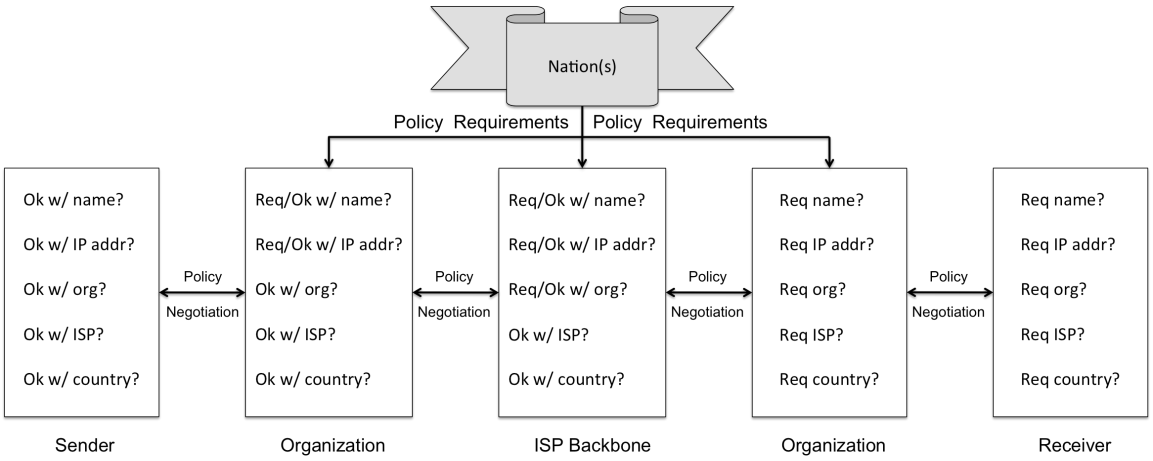

Figure 1. A General Attribution Framework

In this section, we describe our framework in detail. The first three subsections discuss each of the above components of the framework. The final one presents system requirements necessary to support the framework.

4.1 Actors

Our concept of attribution involves an expanded definition that includes interests other than that of the recipient; it encompasses the interests of senders, network perspectives, and other (possibly secondary) requirements. Our definition of attribution, therefore, is much broader in concept, involving multiple parties with multiple intentions, spanning geographic, cultural, social, legal, and national interests. To the best of our knowledge, no one has determined all of the requirements for an attribution system, including who the interested parties are, and what their requirements and incentives are.

We begin by identifying at least nine different entities that have an interest in attribution with respect to a message:

1. The sender of the message;
2. The organization associated with the sender;
3. The governments[2] of the country of the sender;
4. The ISPs over which the message transits;
5. The network backbone providers over whose backbones the message transits;
6. The governments of any intermediate nations through which the message transits;
7. The governments of the country of the recipient;
8. The organization associated with the recipient; and
9. The recipient.

Multiple parties shape whether or how we can show that the characteristics associated with an entity have a particular value. Each of these entities has a distinct and different set of interests in attribution; understanding what attribution really means rests on understanding what these interests are, and under what circumstances the varying interests of different parties can be reconciled, and under what circumstances these varying interests *cannot* be reconciled.

REQUIREMENT 4: The attribution system must include the different actors that have an interest in attribution other than the sender-recipient pair.

4.2 What is being Attributed?

Multiple parties shape whether or how we can show that the characteristics associated with an entity have a particular value. Each of these entities has a distinct and different set of interests in attribution; understanding what attribution really means rests on understanding what these interests are, and under what circumstances the varying interests of different parties can be reconciled, and under what circumstances these varying interests *cannot* be reconciled.

Attribution is thus shaped by three considerations.

First, the "desired attribution" is the product of the interests and capabilities of the different parties involved. The desired attribution depends on three interrelated factors: the desired confidence in the attribution, the nature of the actions for which attribution is desired, and the intended purpose of the attribution. The desired confidence in the attribution is captured in the term "provable attribution" introduced earlier. As to the other two factors, consider the situation where both senders and recipients may want or need absolute non-attribution (e.g., political dissidents operating under repressive regimes). The logic of desired attribution is in a sense circular: the adequate degree of attribution depends on the purpose for which the attribution is intended, which includes the types of responses and actions that might be taken based on the degree of attribution.

Second, the level of certainty associated with showing that the characteristic associated with an entity has a particular value, or one of a particular set of values. This level of certainty – *the attribution*

[2] For example, in the United States, the state and federal governments are different. County, municipality, and other political subdivisions may also have their own interests.

assurance (see earlier discussion) may vary. Today, for instance, authentication for a web site requires less assurance as to the true identity than does authentication for obtaining a passport.

Third, the level of attribution achieved, and the level of attribution desired, may differ. Both the desired and achieved levels of attribution depend on choices made by many different parties involved in the creation, transmission, and receipt of a message. Senders, receivers, and other parties (including network operators but also governments, law enforcement agencies, and companies—but also potentially including other organizations of any sort) all may have distinctly different desires in what they regard as useful attribution; these desires may change depending on the circumstances.

4.2.1. Attribution Agents On the Packet Path

At the end of Section 2.1, we postponed the discussion of the characteristic, and further assumed that the characteristic of attribution would be the identity of the entity. Of course the characteristic can and most likely would include many other types of data – time of transmission, or network data, for instance.

We have further assumed (implicitly) that the characteristic would originate from either the sender or recipient. But the *attribution agent* (the entity to which the characteristic is bound) and the *attribution target* (the entity who receives the attribution) can be other than the sender and recipient. With potentially nine different classes of entities (including sender and recipient) having an interest in message attribution, we have to allow for attribution of characteristics between entities other than the sender and receiver.

Example 12: Jeffrey sends message M to Matt. Along with the attribution binding Jeffrey's machine to the message, so that Matt can attribute M from Jeffrey, Jeffrey's ISP also binds originating time and router data as an attributable characteristic for use by Matt's ISP. Hence there are two sets of attribution taking place with M: the attribution of M from Jeffrey, and the attribution of ISP data from Jeffrey's ISP to Matt's ISP. In this case we have two attribution agents – Jeffrey and Jeffrey's ISP, and two attribution targets, Matt, and Matt's ISP.

Having multiple attribution agents associated with a single packets path raises several issues. Network load is obviously a concern if in practice multiple attribution agents will be active along a packets path. A second issue is whether attributions, say, between Jeffrey and Matt's ISPs, will be visible (or available) to other entities along the packet path, say Matt (or a network backbone provider, or even other ISPs).

Here, attribution privacy becomes a special concern. Will Jeffrey attribution to Matt be visible to intermediate nodes? Might it be an issue if Matt can see the inter-ISP attribution? One possible rule to address this might be to tie access to attribution characteristics to the specific network layer involved; hence Jeffrey's attribution is at Layer 7, while ISP attribution is confined to Layer 3 (hop-to-hop).

REQUIREMENT 5: The attribution framework must allow for attribution to originate and be received by entities along the packet path other than the sender and recipient.

4.3 The Attribution Vector

We use an *attribution vector* to capture the multiple characteristics that are bound to an entity**.** An attribution vector consists of a sequence of pairs. The first element of each pair is a characteristic for which a value is either present or desired. The second element is the value of the characteristic. If the value is not known, the second element is the distinguished symbol ⊥,.

Various types of characteristics will recur when attribution is requested. In practical terms, probably the most common characteristic will be the origin or source of a message to a person or organization. Here, "source" may mean originating user or IP address; it may also mean who originated the information that was sent. Other common characteristics will be the time at which the message was sent, the time at which it entered and exited various networks, the route that the message took (which gives information about who has access to read or alter it), how the message was protected (for example, by encryption or access control bits), and where geographically did the message travel (which may bear on delays or the appropriateness of the mechanism chosen to protect the message).

REQUIREMENT 6: There must be a way to define and then specify the values of the elements of the Attribution Vector

4.4 Attribution Assurance Revisited

Underpinning the values in the attribution vector is the level of assurance of the values. Values supplied by untrusted sources are less credible than values supplied by trusted sources. For example, asking an ISP for assurances that a government intelligence agency did not read messages transiting that ISP would produce assurances of little meaning if the ISP were known to share its data with the government regularly. As already noted, the degree of confidence in attribution depends on its intended use, and possibly on the source of the values that are attributed.

REQUIREMENT 7: The levels of attribution assurance must be specified or determinable by the attribution framework (see Requirement 2)

4.5 Specifying the Sender and Recipient of the Attribution

The origination of the attribution is thus important. Typically one thinks of attribution as relating a packet back to an originating machine, but as discussed earlier this may be insufficient. Desired attribution may be to an individual, or to a role, or an organization, or *a class of individuals*. For instance, it may acceptable to attribute a message from a medical doctor, or a sender over the age of 21, without requiring any further characteristics.

Two other considerations affect how the attribute values are handled. The first is to whom the information is reported. Attribution is traditionally thought of as in the ability to determine, based on the interest of the recipient, where the message came from. But how is attribution handled in instances where (for example) one's spouse is an acceptable attribution recipient, but one's employer is not? More generally, one can consider attribution information as being reported to: 1) the recipient; 2) other entities (e.g., the recipient's spouse); 3) some central authority (e.g., a government or a set of governments) or 4) other intermediate nodes, who either for their own purposes or to pass the information on way find it of value to know what traffic is occurring between two different locations.

REQUIREMENT 8: It must be defined to what entity the attribution is reported or made available.

Note that meeting this requirement has to consider the earlier discussion about the requirements, if any, for *attribution privacy*.

Issue: The last consideration is the characteristic of why the message was sent. Perhaps this is the most challenging information to attribute, but in many situations imaginable, it will be the most important aspect of attribution. An adequate answer however remains an open research question, especially because of the need to examine human motivations. Those are notoriously hard to determine by skilled investigators, let alone by an automated system.

(POTENTIAL) REQUIREMENT 9: There must be a means to define and then specify the characteristics for 'why' the message was sent.

4.6 Reflecting the Interests of Actors other than Sender and Recipient

As noted above, other stakeholders participate in determining type and level of attribution. The ISPs and backbones over which the messages travel move data. They may, or may not, add or delete attribution information:

Example 13: If the originating host's IP address is assigned using the NAT protocol, the firewall (which does the NATing) effectively eliminates the ability to attribute host origin behind the firewall. But the ISP can attribute IP origin to a subnet, here the one with the firewall connected to the ISP. In order to attribute further, the firewall would need to keep a time-stamped log of internal address assignments, and the ISP would need to record the time each packet left the firewall.

This highlights a central issue for ISPs and backbones to provide attribution. What is the financial cost? In particular, ISPs may want to provide attribution services only if they are profitable and the ISP is unlikely to be sued. This balance of profitability and liability is central to the business judgment about whether to provide any service.

Included in the liability issue are cultural and legal constraints. For example, privacy rules in the European Union are considerably more restrictive than those in the United States, so an ISP in the former would be unable to provide the attributions that the latter could provide. In some cases, this

may be a choice. The anonymous remailers are a good example. Cypherpunk type I remailers provide limited non-attribution because, if a list of the pseudonyms and senders are kept, a court order will enable authorities to derive attribution data. But a Cypherpunk type II remailer prevents this by using sophisticated cryptographic and traffic routing and fragmentation techniques.

Organizations are a different matter. As noted earlier, the organizations of interest are the sending organization, the receiving organization, their governments, and the governments of the countries through which the message transits. As an example of the importance of these entities, a message being sent from the United States to Russia over a network that transits North Korea may result in questionable attribution information being added. Thus, the attribution characteristics from intermediate nodes, or that relies on intermediate nodes, is affected by the organizations controlling those nodes.

Other potential issues include hiring and training people to ensure the attribution infrastructure, and other supporting infrastructures, function properly, and that technical and non-technical constraints are met. In addition to the financial burdens of people and processes, the time and other resources required must be considered.

REQUIREMENT 10: There needs to be a means to define, and then specify, the requirements/interests of ISPs and backbones

REQUIREMENT 11: There needs to be a means to define, and then specify, the requirements/interests of other parties

V. A POLICY NEGOTIATION SYSTEM

Directly or indirectly, all stakeholders participate in determining type and level of attribution. To begin consideration of what requirements this creates for the attribution system, let us begin with two broad categories of communications.

5.1 The Need for a Policy Negotiation System

5.1.1. Cooperating senders and receivers

Senders and receivers that co-operate provide attribution capabilities. Consider the case where both sender and recipient agree on a desired level of attribution, as well as specifically to the party to which the attribution applies. The simplest situation is where the sender organization and government are in agreement with this desired level of attribution.

This agreement requires carefully defined and commonly accepted attribution characteristics, and a mechanism for negotiation among all of the parties to ensure agreement on the attributes to be

communicated. So it is in all parties' interests to have a robust system to ensure the agreed upon level of attribution.

Backbones and intermediate nodes, however, have no generic incentive for co-operation. Thus, cooperating senders and receivers have to specify some attributes of the network path (for example, no packets can go through North Korea) to enhance or ensure the required attribution.

Cooperating entities with similar needs create new capabilities: mechanisms for either agreeing in advance on the desired level of attribution and the services needed to support the agreed upon level of attribution, or in having an efficient negotiating system. Furthermore, ideally there would be metrics for the trust placed in backbones and intermediate nodes. A policy based path routing would also be necessary to ensure the paths provided the appropriate support for attribution.

REQUIREMENT 12: There must be a structure for *efficiently* defining policies for the special case of cooperating senders and recipients

5.1.2 Conflicting senders and receivers

Senders and receivers with conflicting attribution needs create choices that either, or both, must make:

Example 14: Political dissidents in repressive regimes provide a scenario that contrasts with that for co-operating senders and receivers. The senders may not (and probably will not) want attribution; whether the recipients would agree to having their receipt of particular packets attributed back to the sender is less clear.

This is a situation in which sending governments (and possibly organizations) want attribution of the sender for repressive political reasons. Recipients, or the international community at large, will probably not want senders to have their messages attributed to them, though this prospect raises the concern that bogus or falsified messages are passed off as legitimate to the recipients.

Furthermore, without the cooperation of sending governments and organizations, creating a policy based routing system will depend on the technical specifications that establishes the policy based trust network, and the extent to which the trust network can in fact be trusted.

In this example, multiple choices exist. Politically dissident senders may simply choose not to use the Internet. Recipients may be less trusting of traffic without sender attribution—for example, how do recipients know that such traffic is not really government sponsored disinformation? Intermediate nodes and backbones may cooperate with the sending government for reasons of their own, thus making the policy based trust network less reliable.

5.2 The Requirements for a Policy Negotiation System

With nine different classes of actors potentially involved in the attribution, typically a policy negotiation will be required in order to establish an agreed upon attribution vector. Such an agreed upon attribution

vector is *a policy contract*. In some cases the negotiations will not succeed; in others, the policy contract will achieve a semi-permanent basis. One can think of policy contract negotiations as a continuum: at one extreme is the oriental bazaar, where everything is constantly negotiated; the other extreme, religious canon, which changes only very slowly if at all. Which structure will predominate we cannot predict; however a policy contract negotiation system must first and always be workable and agreeable to all parties. Given this snap shot of different goals and needs of the different parties with a stake in attribution, having defined who all of the players are and their needs, a full attribution system needs to have several features.

REQUIREMENT 13: A common nomenclature for attribution vectors must be defined: These policy contract elements provide a precise and mutually understood structure including a common language that each involved party can use to define the desired attribution state. The desired attribution state might include the length of the agreement, specified trust levels among network parties (particularly ISPs and backbones), and penalties for non-performance.

REQUIREMENT 14: A system for communicating and negotiating the policy contract must be created. Among and between the different parties this system should be transparent, low cost, and made routine to the extent possible. No system that requires a complex legalistic structure in anything but a few rare cases will work for a commonly accepted attribution framework.

REQUIREMENT 15 and 16: The policy negotiating system should be able to specify and communicate 1) desired attribution states and 2) desired levels of assurance. Satisfying this requirement enables the parties to inform one another in advance of what they require the values of specific attribute characteristics to be in order to accept or reject messages, or continue or terminate policy contract negotiation. At a minimum the senders must be able to specify a level of attribution and the receivers must be able to communicate what levels of attribution it finds acceptable. For example a sender may require that messages not be attributable to its source; the receiver may require full attribution to the source

REQUIREMENT 17: There must be a flexible verification mechanism for ensuring that contracts are performed. Such mechanism will ensure that the entire policy contract negotiation mechanism is enforceable. The verification mechanism needs to provide consequences for those who follow, and fail to follow, negotiated contracts. For example, it might publicly note those who honor policy contracts and those who do not, by using a reputation-based system; or, it may impose a punishment system for violating agreed upon policy contracts, up to and including ostracizing those who breach them.

Note importantly that policy negotiations themselves cannot violate existing policies:

Example 15: A sender may already have as its policy that its identity never be attributable. Whether a negotiation can succeed under *existing* policies is a question of some import, especially because those policies may not be known when the negotiation starts. One possibility to ameliorate this is to provide a trusted storage mechanism for existing policies, which specify the framework for any further

negotiations, or identify specific types of policy negotiations that may take place between either wholly or partially anonymous parties.

Example 16: As another example of the complexity of policy negotiations, a government web site might require attribution to the user level, but be willing to negotiate down to just an IP address should the user prefer not to provide his personal identity.

It is important that the policy system avoid allowing unwanted accidental outcomes – in other words, situations where the attribution 'agreed to' by the entities is not in fact what is desired. In some cases, accidental outcomes could have serious implications as in the following example:

Example 17: A dissident web site needs to advertise its policy of not accepting any forms of attribution *before* a prospective user accidentally provides it.

We therefore need to determine what policies might be required and how they might be made known to other participants. Mechanisms for advertising policies need to be devised, along with an examination of where policies will reside (for example, in addition to policies at the end points, intermediate routers may also have policies that all transiting traffic must honor).

Example 18: Looking at a different application is when the "negotiation" that takes place between a recipient with a telephone blocking calls that suppress the caller ID, and a caller whose telephone does not transmit the caller ID (clearly requiring some other mechanism to initiate communication, or simply the sender determining that communication is not possible).

Finally, the mechanisms must be available to non-participants who wish to join the circle of negotiation in order to communicate with entities that require policy contracts.

In many cases, one party may act as a representative for a class of parties to determine a generic policy contract. This is akin to "class action lawsuits," in which a set of actors with a common interest authorize one actor to negotiate on their behalf. In this case, the policy negotiation mechanisms must enable the binding of all parties, not just the negotiator, to the contract.

This leads to some specific system constraints that support policy negotiation.

REQUIREMENT 18: A trust network must be defined that enables actors to trust that other actors, and the network, will honor their commitments as negotiated in the policy contract. Networks cannot tag or alter packets of their own accord[3]; some entity must set them to do so. Thus, signers of a policy contract must have some measure of trust in the other actors to provide attribute values, and to provide *acceptably accurate* values. This trust system might be tied to the verification system mentioned above, and function much as a reputation system would.

[3] Excluding errors

REQUIREMENT 19: A policy-based routing mechanism must be defined to ensure that messages traverse networks and midpoints with appropriate attribution mechanisms and levels of trust. This is particularly important if messages are to be routed dynamically (as in today's Internet). Since an attribution 'wrapper' around a packet while technically conceivable would be difficult/impossible to put into practice, the intermediate nodes can alter the attribute vector or add attribute data of their own; hence trust in attribution will be based in part on routing.Unless the actors do not care whether the attribution changes in transit the path that the message takes affects both the values in the attribute vector and the level of assurance of that vector (including the values).

5.3 Automated Negotiation

This subsection will look at some automated policy negotiation systems such as SCENS that might, or might not, work here. We will examine whether each will, and if not what technology needs to exist for automation to succeed, if it can.

VI. GOVERNANCE OF THE ATTRIBUTION SYSTEM

A number of policy and social issues arise in the creation of an attribution system. How these issues are resolved may not be the choice of any system designer; rather, it may be, as is the case in some other infrastructures (the Internet today being a notable example) governance issues will be resolved – or not – in an evolutionary, if not to say, 'messy' fashion. Nonetheless, there are key governance issues which, while not exactly requirements, will have to be managed over time, if not ever solved. Three are highlighted here;

- Who makes or how are key attribution framework decisions (e.g., dispute resolution) made?
- How might economic factors shape the form of the attribution system?
- What are some of the social considerations that affect the way attributions are made?

The following sections address these issues.

6.1 Who Makes the Decisions?

One aspect of this foundational governance issue is the creation of a "superuser" or "Administrator" for the attribution system, in which one privileged user can override normal user controls. A likely future issue will be defining the role of central authorities (governments, or network authorities, or both) from overriding the policy-based routing and and trust networks under defined circumstances. Traditionally, this mechanism s is used to provide an escape to correct severe problems or failures. In high assurance systems, this omnipotent role is partitioned into a set of less powerful roles. What powers such a role should have in the systems implementing the policy negotiation, or indeed whether such a role should exist, is an open question. In theory, a superuser should not be needed because the actors in the negotiation can simply decide no agreement is possible. But in practice, other authorities (such as

governments) may require such a role for non-technical reasons as for instance when law enforcement requires attribution for a certain set of messages between two parties (one or the other party may not voluntarily agree to this!). If so, how such a role would be implemented across multiple jurisdictions is a difficult question, especially when the jurisdictions involved are those of different nations.

Other issues include the extent to which common protocols to implement the policy negotiation system must be adopted. This depends in part on the goals of the system. If attribution is to be ubiquitous, then common protocols (or at least interoperable protocols) must be adopted. Alternatively, several policy negotiation systems might exist, each supporting different types of attribute vectors or different levels of assurance for attribute vectors; in this case, the ability to map goals from one system to the other, and to create translation mechanisms to allow the resprctive protocols to interoperate, define the extent to which attribution information and trust may be shared.

In fact, none of these issues is unique to networked systems; the world of negotiating structures and mechanisms is well established in the non-technical world, and many mechanisms exist in the technical world to support negotiations. All of these issues have been resolved or at least managed in various ways in the physical world (including a realization that, in some cases, negotiations are not feasible). These issues facing an attribution system in turn raises a number of further issues, including:

- What is "adequate" attribution or authentication? This is of course a governance question—who decides? And how might governance goals reflect changing needs of users/administrative domains?
- To whom are the attribution values accessible under select circumstances? Will there be negotiations required if multiple central authorities are involved in an attribution; for example, will multi-jurisdictional cooperation depend upon certain limitations of the form of response allowed?
- Under what circumstances can attribution be "undone"? For instance, under perfect attribution, a central authority could tell networks (intermediate nodes) not to record or tag packets.
- How should conflicts and ambiguities among users/administrative domains be handled? For example, attribution may be desirable under some circumstances (cyber attack and crime) and undesirable under others (political free speech, and possibly of "whistleblowers"). Entities/ADs may have different, conflicting goals here, and hence the success of a governance system in resolving such conflicts will form part of its evaluation (though the metrics will need to be developed). Of note the political and cultural aspects of attribution, where we from a Western culture assume that the ability to visit a dissident web site is good, whereas the government of other countries would obviously disagree with our particular beliefs and morals. In some cases, notably when a whistleblower reveals information his or her organization would prefer to keep secret, the organization will want full attribution whereas doing so would be inimical to the whistleblower.
- How should special (e.g., national defense) needs be handled? There are cases where there may be a requirement for false attribution, such as in the example of when an intelligence agent is undercover and visiting a terrorist website, and where the terrorist website might require

attribution. Should this capability be provided to, say, intelligence agencies in all circumstances? If not, under what circumstances should it be provided, and who should determine whether those circumstances hold?

6.2 Economics Will Shape the Attribution System

Most likely economic considerations will shape the incentives and structures for a full attribution system in several different ways.

First, there is a substantial body of work demonstrating that trust and privacy have a real economic value. We have not proposed a mechanism for monetizing this economic value, but creating a market for attribution and non-attribution among the nine sets of participants appears to be an attractive option. Senders, receivers, and the intermediate organizations could make side payments in order to achieve the desired attribution outcome. The total (dollar) size of this market, how such markets would function is actually buying or selling contracts for 'attribution' (or variants of attribution, as earlier discussed), and how such markets would be governed, are speculative right now, but such markets appear to be conceptually attractive. The real-life examples of robust markets in derivatives, options, and other esoteric 'non-material' products suggest that a market for attribution trading might be feasible. Even if such markets are, in economic terms, 'imperfect', so that the value of trust and privacy is not (fully) monetized, a functioning market nonetheless would contribute to the overall economic welfare (in economist's terms, either in terms of the consumer or producer surplus).

Second, backbones and intermediate nodes face a couple of different economic models for their businesses. For example, intermediate nodes and backbones could position themselves as the most trusted intermediate carriage points for traffic with attribution or non-attribution requirements. In this case, the rationale is that by being highly trusted these carriers would obtain more traffic (but this assumes that the market for attribution will in fact be significant). Alternatively, nodes could adopt a low cost strategy—make no guarantees as to the validity of the traffic crossing these nodes, but count on transmitting significant traffic at a low cost. A more venial instance would be for nodes to accept side payments (from governments or organizations) in order to corrupt or monitor their traffic, without the knowledge of other attribution system participants.

Governments and organizations also have to make choices as to how they are positioned in providing a trusted range of attribution choices. To cite a banking analogy, at one end of the spectrum are the trusted Swiss; at the other end would be countries like Nigeria.

Our intuition is that the economic flows from a full attribution system will be considerable (though we have no evidence to support this assertion), and that a variety of business models will emerge variously trading off trust, traffic volume, cost and even side payments from other parties.

Policy choices may shape the ultimate network economics. By treaty, international telephony provides payments to less developed countries to support their connection to the multinational network (in total, such reverse payments are on the order of 8-10 billion US per year). The Internet has no such structure

of reverse payments, but such a system might be a powerful incentive for select countries to provide and participate in a trusted attribution system. This payment structure deserves careful analysis.

Finally (and related to the previous paragraph) is the question of "who pays." The attribution system as outlined in this paper would require significant investment in multilateral capabilities that do not now exist. These include:

- A common multilateral policy framework to formalize the cooperation, definitions, and collaborations necessary for attribution across administrative, jurisdictional, and national boundaries;
- Technical cooperation far exceeding the agreements in principle now extant. Such cooperation would fill important gaps, such as research and recommending the best attribution techniques, and providing on-going support for a multilateral attribution capability;
- Negotiating structures (not just for senders and receivers, but all nine sets of parties involved) with defined terms for levels of attribution and non-attribution to be associated with each message; and
- Policy based trusted network routing across backbones/nodes. Ideally a formalized metric for trustworthiness would be developed and used as the basis for routing decisions.

All of these initiatives are a necessary part of the attribution system we have outlined. While all would benefit, as is typical of network transformations appears to be little incentive for any single party or group of parties to fund these initiatives. As we have noted, there are successful multilateral frameworks that have been developed to address needs like non-proliferation and weapons limitations. We are confident, therefore, that an acceptable system for funding this attribution overhead is quite doable.

6.3 Social Factors May Shape the Attribution System

The attribution policies discussed so far create an interesting, and yet realistic, dichotomy. Consider the attribution policy of first origin, that is, the ultimate source.

This policy states that the network operators can trace coordinated entities back to their origin. The utility of this policy arises from distributed DDoS attacks, or botnets, in which the immediate origins of the messages are known (the bots); the policy requires the ability to trace back to the distribution points, or distributors, of the bots. In the context of tracing network attacks, the first origin policy is not merely reasonable; it is salutary, because it minimizes disruption and suspicion of those unwitting people and systems on which the botnet entities run.

Now, consider the same policy in a political context. A nation with repressive political policies discovers a large number of messages that poke fun at the government. The first origin policy allows the government to trace back to determine the origin of these co-ordinated entities (one or more messages). The ability for the dissidents (or ordinary citizens) to criticize their government anonymously no longer exists.

This leads one to ask the purpose of attribution. Attribution, or rather the lack of attribution, provides the ability to send messages without fear that the entities involved can be identified. Differing levels and types of attribution modify the level of fear, and the ability to send such messages, in various ways.

The ability to conceal the origin of messages affords the sender protection from reprisals. The example using political dissidents is one context in which this ability is critical. Another example is whistle blowing, in which a subordinate reports actions of a superior (or an equal) to an external authority, such as the press or a regulatory or law enforcement agency. Extending this to an agency or country, the ability to deny attribution allows an attacker to place the target in a state of confusion, a tactic of warfare encouraged by Sun-Tzu, among others.

This ability also enables one to protect privacy. Called “the most comprehensive of rights and the right most valued by civilized men” [Warren & Brandeis], the right to be ”let alone” enables one to live one’s life without interference and without having to account for one’s actions. As an example of the value of such privacy, consider someone who wants to learn to use the Python programming language. He goes first to the web site http://www.python.com. The pornographic images on that site indicate it is not the site where one may download the language interpreter, so he tries http://www.python.org, which is the correct site. But anyone observing his activities would see he visited a pornographic web site, and from that could (erroneously) conclude he was downloading pornography.

With privacy comes power. A lack of attribution enables entities to avoid taking responsibility for their messages. For example, experience with the anonymity that the Internet affords shows that it prevents those who are the targets of slanderous communication from identifying the sources, and taking legal or other actions to protect themselves.

The sources here may include the government. Franz Kafka’s book ”The Trial” makes this point eloquently, by describing a trial in which the protagonist is tried for something (he is never told what) by a court (he is never told who), and subsequently convicted and executed. In many countries, people are tried (either in a court of law or in the court of public opinion) without being told who their accusers are. So the lack of attribution that protects the individual also can harm the individual.

Further, the ability to trace messages enhances the ability to detect attacks at the non-cyber level, ranging from individual threats (for example, harassment) to societal threats (for example, terrorism and warfare). Thus, this point of view stems from a belief that providing attribution encourages social order and protects both individuals and society.

There is no right answer to the level of attribution that should be provided. This is a policy issue that must be decided somehow, either by a deliberate crafting of policy or by an acceptance of the existence of tools and services that can provide varying degrees of attribution.

Ultimately, there may be several Internets or slices . . ., each with a different level of attribution, and people who desire disparate levels may simply be unable to communicate. While disquieting, this

mimics the non-cyber world perfectly. Two people may talk, but one may not believe the other's claims because the attribution of those claims is insufficient for the skeptic's purpose. That the speaker cannot provide the level of attribution that the listener desires interferes with communication, and in some cases simply cannot be overcome. So in this way the use of attribution in cyberspace has the same effects as the use of attribution in realspace.

VII. Summary and Conclusion

(To come)