An End-to-End Argument for Network Layer, Inter-Domain Access Controls
TR-90-13

Deborah Estrin   Gene Tsudik

Computer Science Department
University of Southern California
Los Angeles, CA 90089

To appear in *Journal of Internetworking: Research and Experience*, 1991.

An End-to-End Argument for Network Layer, Inter-Domain Access Controls

Deborah Estrin   Gene Tsudik

Computer Networks and Distributed Systems Laboratory
Computer Science Department
University of Southern California
Los Angeles, California 90089-0782
estrin@usc.edu   *tsudik@usc.edu*

April 25, 1991

Abstract

In the absence of special mechanisms, network interconnection using DOD or OSI Internet Protocols achieve full connectivity among internet components. However, independent administrative domains (ADs) should be able to interconnect without exposing their internal resources to unrestricted access. Moreover, ADs should be able to protect incoming and outgoing data by specifying or constraining the ADs to, and through, which the data travels. Given these requirements, what layer is most appropriate for the placement of access controls?

In this paper we propose a framework for placement and design of access control mechanisms in an environment of interconnected ADs. To the extent network resources require protection, the highest relevant endpoint is the network router and associated packet forwarding and routing protocols. In this sense the *end-to-end* argument supports implementing these controls at the network layer.

We describe the security services that are required by interconnected ADs and address the design choices of enforcement location, protocol, and granularity. We conclude by addressing how the various layers and mechanisms interact.

1 Introduction

In the environment of interconnected Administrative Domains, access to network resources is an issue of growing concern. As described in [9, 13, 16, 7], an Administrative Domain (AD) is a set of resources (network links, routers, bridges and end-systems) under the control of a single administrative authority. In the absence of special mechanisms, network interconnection using DOD IP[20] or OSI[11] internetworking protocols attempts to achieve full connectivity. However, ADs should be able to interconnect without exposing their internal resources to unrestricted access[4, 6]. Moreover, ADs should be able to control the incoming and outgoing data by specifying or constraining the ADs to, and through, which the data travels[5].

In this paper we address the subject of access control in the interconnected AD environment. We begin in Section 2 by defining the three objects of access control that exist in a multi-AD internet – end-systems, network-layer resources, and inter-AD routes. We address the well-known *end-to-end* argument in proposing controls in both transport and network layer protocols[23]. Section 3 describes the security services needed to address each of these requirements and the corresponding design choices of enforcement location, protocol, and granularity. We conclude in Section 4 by addressing how the various layers and mechanisms interact.

2 Control Requirements

An internetwork is composed of different types of ADs which require different types of control. We distinguish among two dominant types of ADs: *stub* and *transit*. A *stub* AD is an AD that provides communication facilities for its constituent end-systems and does not carry transit traffic for other ADs (e.g., a campus network). A *transit* AD's primary function is to provide transit services for other ADs (e.g., long-haul backbones and regional networks). There are also *hybrid* ADs that support access to their constituent end-systems, as well as limited transit to other ADs.

This section begins with a brief discussion of end-system control in an inter-AD context. We then argue for controlling access to AD network resources independently of the end-system access control.

2.1 End-systems and Applications

End-system security is a requirement for all stub ADs. Much previous work has addressed the design of secure applications, operating systems, as well as the adaptation of secure systems to a network context [28, 17, 29, 28, 14, 6].

Modern distributed operating systems, e.g., Andrew [25] and Amoeba [18], illustrate methods for efficient implementation of security features in a distributed computing environment with high availability of services. In the realm of secure applications, Privacy-Enhanced Electronic Mail [17], for example, provides valuable insights pertaining to the implementation of security services in a large-scale distributed environment with a very large and volatile user population.

In this paper we emphasize end-system and application-level controls as a point of comparison for the other controls addressed, and comment on the division of labor between end-system and network controls. But first, we briefly discuss how inter-AD connections influence the requirements for end-system controls themselves.

Inter-AD connections impact ADs that have relatively open internal computing environments more so than ADs with closed, protective internal environments. For an AD that employs rigorous security mechanisms on *all* end-systems, introduction of an inter-AD link need not induce significant modification of the end-systems themselves. However, many ADs employ a more *laissez-faire* approach to security – implementing stringent controls on only a small subset of critical systems and leaving the balance of end-systems relatively vulnerable to intra-AD access. This should not be viewed as a criticism or failing of intra-AD security. Rather, to the extent that controls may inhibit internal resource sharing, it may be in the organization's interest to maintain a relatively open environment. Introducing additional controls throughout an AD is often impractical because of the large number of internal end-systems and the subsequent difficulties in verifying the correct operation of these controls. Nevertheless, introduction of inter-AD links may require some reassessment of this approach.

In addition, some communication-oriented applications will require special consideration. Among them are: electronic mail, video conferencing, file sharing and process migration. Each demands a unique combination of security-related services. In video conferencing, for example, privacy is often important while data integrity is somewhat secondary; also, bandwidth availability is critical. Hence, denial of service (malicious or otherwise) is of significant concern.

2.2 Network Resources

Many discussions of *network security* are actually discussions of end-system protection in a network environment, e.g., [27, 19, 15, 14, 6, 29]. While this is an important consideration, we claim that it is not adequate in the multi-AD context. For both stub and transit ADs, there are valuable network resources that are also the object of policy.

This is in agreement with a well-known design principle, the *end-to-end* argument[23]. It states that the placement of controls should be in the highest protocol layer at the end-points of communication. From a functional standpoint, features such as reliability and security must take place in the highest layers if they are to cover all sources of vulnerability. Therefore, the argument is that lower-layer efforts are always redundant and should be implemented only to the extent that they improve efficiency. In the case of security, the argument suggests that end-system resources are best protected by the end-systems themselves, e.g., security services should be provided in the transport layer as opposed to the network or data link layer [29]. We observe, however, that in the sense of the *end-to-end* argument, network resources are *endpoints* to the extent they require protection in their own right. Therefore, it is imperative to address the protection of network resource in addition to the more traditional end-system protection. This implies the need for controls at the AD exit and entry points.

If control is left to the end-systems, valuable stub-AD network resources may be consumed by unauthorized traffic. Rejecting packets at the end-system is *too late* from the perspective of network resource usage. Moreover, unrestricted network access increases the vulnerability of ADs to denial of service attacks in the form of packet storms. In other words, the network interfaces of end-systems are themselves network resources and should be subject to access control (which can only be achieved below the transport layer). Similarly, some ADs have no relevant end-systems (they provide transit services only) and therefore must implement desired controls in the network-layer routing protocol. Since routing is a network layer function, these controls must also involve network level entities and can not be left to transport session endpoints.

Network interconnection is typically done at the network layer for reasons of transparency, flexibility and performance. Consequently, transport and higher layers are generally not well-suited for the implementation of network resource controls. Moreover, lower layers (physical and data-link) do not provide access to subject and object information necessary to make access control decisions with respect to network resources. This leads to our assertion that the network layer is most appropriate for the implementation of network resource controls.

In summary, to the extent network resources require protection, the highest relevant endpoint is the network router and associated packet forwarding and routing protocols. In this sense, the *end-to-end* argument supports implementing these controls at the network layer.

2.2.1 AD Boundaries

An AD represents a set of resources that are governed by common policies. We argue that the enforcement of policies pertaining to inter-AD communication is best carried out at the AD boundaries.[1] Throughout our discussion we will exploit the AD abstraction to decouple requirements and mechanisms for intra-AD communication from those for inter-AD communication. In particular, we assume network layer enforcement only in those routers that connect the AD to other ADs, i.e., border routers (also referred to as Policy Gateways [16]).

Another motivating factor for implementing controls at border routers is the trend towards commercialization of portions of the Internet. Some network resources may be charged for on a usage-sensitive basis. This opens new incentives and opportunities for fraud. Consequently, border routers must be equipped to detect such abuse before valuable network resources are consumed and charges are accrued.

In the following sections, we address such security concerns in the context of both stub and transit ADs.

2.2.2 Stub ADs

Stub ADs need to protect the integrity of their internal network in the presence of inter-AD connections. Network resources that may require protection include links, bridges, routers, and end-system network interfaces. In the following paragraphs we further justify the need to control access to communication resources themselves, and the need to include end-system network interfaces among the resources being protected.

Most internal networks are implemented without explicit access controls in the nodes or routers. In the context of intra-AD use, limitations on access to private information or resources applies primarily to end-systems. Communication is most often treated as a *free good* within organizations. It is neither charged for, nor controlled on a usage-sensitive basis. Most often, internal network access is unrestricted if a user has legitimate end-system access. Even within the AD this can

[1]Policies that apply to intra-AD communication can be enforced by each AD independently.

lead to undesired dependencies among hosts, departments, etc; for example, misbehaving hosts generating broadcast storms, or poor transport protocol implementations degrading performance of a bottleneck resource. However, in most cases, the existence of a common administrative umbrella alleviates the need for intra-AD control of communication resource usage.

When an inter-AD connection is first established, it may violate some of the assumptions under which these intra-AD protection decisions were made. There is no longer a common administration, nor the common organizational goals that can be assumed in the intra-AD context. In addition, there is no control over implementation and configuration of end-systems and routers in other ADs. The number and nature of the potential users increases qualitatively in the presence of an inter-AD link. In this context, the communication resources can no longer be considered *free* to all potential users. For this reason, stub AD network resources themselves must be considered as objects of access control policy in a multi-AD internet.

The external connection in a stub AD usually exists for the purposes of a small subset of internal end-systems[4]. It may be undesirable to expose all internal end-systems. Furthermore, as discussed in section 2.1, it is impractical to assume that all end-systems in every AD will implement adequate defenses to be considered secure in the face of a greatly expanded and diversified user community (e.g., a global internet)[2]. Therefore, ADs require mechanisms that allow them to designate those end-systems that will be reachable from the outside, and those that will not. Similarly, the AD may wish to restrict origination of outgoing traffic to those end-systems that are explicitly authorized, e.g., to restrict the usage of pay-per-packet transit services or to reduce the risk of undesired information exportation. Hereafter, we refer to the externally-accessible (reachable) end-systems as *equipped*, and the rest as *unequipped* end-systems.

Access control to unequipped end-systems is treated as a network level access control problem. In other words, the end-system network interfaces are viewed as network resources that require protection. By controlling access to end-system interfaces, ADs can also address the susceptibility of all end-systems to denial of service attacks through flooding [6].

Finally, even communication with equipped end-systems must be controlled at the network layer. Each time an unauthorized external user attempts to communicate with an equipped end-system, stub AD network resources are consumed. Even if the end-system knows to reject the transaction, it is too late with respect to the expended network resources.

In summary, if a stub AD wishes to control usage of its internal network resources (links, bridges, routers, and end-system network interfaces) it can not rely solely on the protection mechanisms in end-systems.

2.2.3 Transit ADs

Transit ADs are concerned with controlling usage of, and access to, their internal routers and links. Their policies may be based upon the source or destination AD, the previous or next hop AD, or other characteristics such as user classes, charge codes, or type of service[5]. Transit resources may be billed on a usage-sensitive basis. In addition, service quality is dependent upon adequate

[2]We consider physical isolation of reachable end-systems from strictly internal end-systems to be an overly restrictive solution. For most environments, it would infringe on internal communication and integration to an unacceptable extent.

capacity to meet demand. Consequently, control of access to these resources is critical to their operation.

As with stub ADs, transit resource protection can not be left to end-systems. By the time traffic reaches the end-systems, the communication resources would have been used. Moreover, in the case of transit ADs, the destination end-system is not within the particular transit AD's administrative control and can not be expected to recover the transit AD's costs. Furthermore, transit ADs should not rely on stub ADs to enforce transit policies; this represents an excessive compromise of transit-AD autonomy. Even if transit ADs are open to all paying customers, they need to monitor and charge for traffic. Monitoring and charging are just different types of network layer control mechanisms. In any case, network layer controls are needed at the boundaries of transit ADs to protect against unauthorized resource usage and denial of service attacks.

2.3 Route selection

The last policy requirement faced in the multi-AD context is control of route selection. Different routes have different characteristics. For example, some routes may travel via ADs that charge on a usage sensitive basis. Other ADs may be avoided because they are not trusted. Stub ADs wish to express policies regarding the particulars of their traffic flows, i.e., where a packet can be sent, how it should get there, and which internal systems can originate it. Similarly, some ADs will want to enforce policies regarding the routes taken by incoming traffic. Control of transit AD network resources and route selection are similar to the extent that they both restrict access to transit resources. However, route selection is based on preferences of stub ADs, whereas transit ADs protect their resources on the basis of local policies.

Therefore, a stub AD may need to control access to inter-AD routes according to their characteristics. Not all internal traffic sources will be given external access. Even those sources allowed external access may not be permitted to use routes that incur usage-sensitive charges. Still others may be permitted to transmit only over routes composed of highly trusted ADs. Some AD destinations may even be considered off limits to all outgoing traffic. These policies can not be enforced by transport layer protocols as routing is transparent to them. Hence, enforcement must take place at the level of internetwork routing and packet forwarding protocols.

3 Design Choices

This section addresses the design of enforcement mechanisms to meet all three control requirements – end-system, network resource, and route selection. It begins with the discussion of applicable security services and proceeds to discuss appropriate enforcement locations, enforcement protocols, and granularity of principals. In the final subsection the object granularity and mode of enforcement are discussed. The results of our discussion are summarized in Table 2.

3.1 Security Services

The OSI model specifies fourteen security services [12]. (They are summarized in Table 1). These services, however, are not uniformly applicable to all types of objects being restricted by policy:

No.	ISO Security Service	Applicability
1	Peer Entity Authentication	-
2	Data Origin Authentication	*
3	Access Control Service	*
4	Connection Confidentiality	-
5	Connectionless Confidentiality	-
6	Selective Field Confidentiality	-
7	Traffic Flow Confidentiality	?
8	Connection Integrity with Recovery	-
9	Connection Integrity without Recovery	-
10	Selective Field Connection Integrity	-
11	Connectionless Integrity	*
12	Selective Field Connectionless Integrity	-
13	Non-repudiation at Origin	*
14	Non-repudiation at Delivery	?

Legend:

* – applicable
- – not applicable
? – potentially applicable

Table 1: Applicability of ISO Security Services

1. End-systems
   Due to the variety of applications and services provided, end-systems are subject to a wide range of potential security threats. To counter these threats, end-systems may require all or most of the (fourteen) OSI security services.

2. Network resources

   (a) Stub ADs
   In order to prevent unauthorized access to their internal network resources (as well as unauthorized export of traffic across AD boundaries), stub ADs need to enforce access control policies. This entails the authentication of principals (peers) involved. Furthermore, since a compromised communication channel between an authorized pair of principals can lead to improper usage of stub AD's internal network resources (i.e., compromised data still consumes stub AD network resources *en route*), data origin authentication and data integrity must be maintained as well. Non-repudiation of origin may be needed where the AD accounts for use of inter-AD resources. On the other hand, confidentiality of end-to-end data is logically a function of the end-to-end protocols and systems.

   (b) Transit ADs
   Because transit AD network resources are similar to those of stub ADs (with the exception of end-systems) they are subject to the same security threats, and, hence, require much the same security services. As discussed later, differences arise in other characteristics of policy enforcement.

3. The route selection policies can be thought of as access control restrictions to internetwork routes. This requires authentication of the principals requesting access to routes. Furthermore, it also requires authentication and integrity of routing information provided by the transit ADs (because this routing information is used to associate cost and security characteristics with computed routes).

In summary, we are primarily concerned with five security services – access control, peer entity authentication, data origin authentication, data integrity and non-repudiation. The remaining services may be of concern to some end-systems but, in general, are not affected by the particular issue of AD interconnection. Not surprisingly, the Standard for Inter-operable LAN Security (SILS) recommends attention to these same five security services[10]. Moreover, the SILS document points out the inter-dependencies among these five services: access control on authentication and integrity, authentication on integrity, and non-repudiation on authentication and integrity.

3.2 Enforcement Location

One of the most critical decisions in the design of enforcement mechanisms is their physical location. In Section 2.2 we justified the placement of controls only in those routers that act as points of connections to other ADs, i.e., the border routers.

In determining the appropriate locations for the enforcement of specific policies, we based our decisions on the principle that, ideally, unauthorized resource usage attempts should be detected before any resources have been consumed, i.e., at the earliest possible point.

1. **End-system resources**: equipped end-systems can be controlled by mechanisms in the transport and application layers. Controls for unequipped end-systems are placed with the stub AD's network resource controls (see next item). This is because no external traffic whatsoever should be allowed to reach these end-systems as they are assumed to be unprotected.
2. **Network resources**: Stub and transit AD network resources are both controlled by mechanisms located in the border routers of the ADs. Border routers may work in conjunction with servers (e.g., authentication and access control servers, policy servers, and route servers) that will also enforce policy. Such servers are needed to offload time and space consuming functions for performance-critical routers, and as a point of coordination and integration of policy (see Section 4).
3. **Route selection**: Stub ADs control access to routes within the route servers that compute external routes and within border routers that may validate traffic.

3.3 Enforcement protocol

Closely related to enforcement location is enforcement protocol. We consider mechanisms in the end-system transport (and higher layer) protocols, the internetwork packet-forwarding protocol, and the internetwork routing protocol. We distinguish between mechanisms implemented in the intra-AD protocols, and those required in the border router protocols only.

1. End-systems can protect themselves at any (or all) layers in the protocol hierarchy (transport layer being the most applicable). Of course, this only holds for the reachable end-systems that are sufficiently equipped.

2. (a) The unequipped end-systems and the rest of the stub AD network resources are protected by the border routers as discussed above. Since ADs interconnect at the network layer and the **network-layer packet-forwarding** protocol is the highest layer with respect to these resources, it is the most appropriate protocol for the enforcement of access restrictions to unequipped end-systems.

   (b) In order to protect network resources of transit ADs, access controls must also be incorporated into the **network-layer routing protocol**. This is necessary because, unlike stub AD policies, transit AD policies may prevent communication even when a viable route exists. If a stub AD's policy disallows communication, the user may be inconvenienced. However, the policy is having its desired affect. On the other hand, transit AD policies have more far-reaching impact. The shortest route computed by a traditional routing algorithm may not be usable by a particular source due to a policy of one of the transit ADs. Without access to transit policy information, the routing protocol has no means of finding an alternative, perhaps *longer*, route for that source.

   As a result, transit ADs can not simply enforce policy restrictions on a unilateral basis at packet forwarding time. Instead, policies pertaining to transit AD network resources must be either implicit in the topology of an internetwork, or advertised to the anticipated resource users as part of the network-layer routing protocol.[3]

3. Route selection policies are enforced at both border routers and route servers. In route servers, the enforcement protocol is the **routing protocol** application that computes internetwork routes and distributes them to appropriate end-systems. In addition, Border routers have to validate route selection made by the route servers at packet forwarding time. This latter function needs to take place in the network-layer packet-forwarding protocol.

3.4 Principal Granularity

We refer to the subject of a security policy as a *principal* [24] (i.e., a principal is permitted to access, or is restricted from accessing a particular object). Policies may be applied to ADs as a whole, to user classes that are location-independent, to particular end-systems within ADs, or to particular users or user processes. The coarser grain policies (i.e, those based on AD or user class) are easier to manage but, by definition, less precise. Of these policies, ones based on user classes are more difficult to implement, but offer flexibility of grouping users independent of the physical location. The difficulty arises from having to bind a user or user class to lower-level units of communication, i.e., packets. Whereas packets routinely include end-system addresses that can be mapped more easily into AD addresses through existing mechanisms.

1. For policy enforcement at the end-system level, the choice of principal granularity depends on the protocol layer. Since end-systems can implement controls at several protocol layers, principals of different granularity can be specified. For example, end-systems at the transport layer and application protocols or users at higher layers.

[3] In [1] we argue that it is difficult to reflect a rich set of policies in the topology when the internet is of this scale and heterogeneity.

2. The ends-points of inter-AD communication with respect to stub AD network resources are the end-point stub ADs This implies that the principal granularity for stub AD access control should be at the level of ADs or user classes. However, as mentioned before, network resources of stub ADs include the *unequipped* end-systems. To preclude all access to these end-systems, it is necessary to discriminate incoming traffic on the basis of the end-system destination address. Therefore, principal granularity for stub AD network resources may be end-system as well as AD.

3. Since transit ADs provide communication facilities for large numbers of stub ADs, it is impractical for their resource usage policy to specify fine-grained principals such as end-systems.[4] Therefore, the subjects of transit resource policy enforcement are expected to be ADs and user classes.

4. For route selection policies, we must consider the granularities of the two end-point (source and destination) principals. They may be the same (e.g., a single route between a pair of ADs) or they may differ, e.g., a policy may prescribe a specific route for a given {*end-system,destination-AD*} pair. It is expected that most route selection policies will restrict based on source AD (or user class) and destination AD.

3.5 Communication Granularity and Enforcement Mode

Policies may be applied to communication units of different granularity. In particular, there are tradeoffs associated with implementing controls per packet, per end-system association (i.e., connection), and per AD association. We also consider the appropriateness of using a priori (preventive) or post facto (accounting) detection, of unauthorized resource use, in the context of different policy objects and security services.

1. At the end-systems, the granularity of controlled communication units depends on their reachability. For the reachable end-systems, granularity depends upon the particular enforcement protocol. For example, per-packet at the network layer, and per end-system association at the transport layer. Regardless of the protocol, the mode of enforcement has to be preventive (at real-time) since end-system disruption can not usually be tolerated.

   Access control for unequipped end-systems has to be done on a per packet preventive basis in border routers. This is because even a single packet can disrupt these unprotected end-systems.

2. It is less straight forward to determine the appropriate basis for protection of other stub AD network resources. If reachable end-systems implement controls, and external traffic to unequipped end-systems is prevented at AD boundaries, the only potential for unauthorized use of stub AD network resources is the compromised traffic addressed to reachable end-systems. Such traffic consumes network resources *en route*[5]. However, authenticating the origin and checking data integrity of every incoming packet may be expensive in terms of performance and implementation. Thus, some ADs may elect to allow for some unauthorized resource usage of this sort in return for faster packet switching in border routers, while others may go to the trouble of scrutinizing every packet to assure non-interference of compromised

[4] Transit ADs are also more likely to bundle usage charges on the basis of ADs rather than end-systems.

[5] However, it poses no direct threat to end-systems, as they are sufficiently protected.

traffic. In summary, for network resources other than end-systems, both real-time and post facto (accounting-based) detection methods can be used.

3. For transit AD network resources, the communication unit granularity is dependent upon the particular security service. It is impractical to control access on the basis of packets or end-system associations due to potentially large numbers of them. It is more manageable to apply enforcement to stub AD associations. Because AD-pair associations are coarse-grained (i.e., can encompass large volumes of packet traffic), establishment of such associations may be verified and validated by intervening transit AD. However, due to the potential risks involved, this type of enforcement should be addressed using preventive methods.

   Once a stub AD-pair association is validated by a transit AD, individual data packets must be associated with a particular AD-pair, e.g., to pass charges appropriately. The implication is that at least some enforcement must take place on per-packet basis. One possible scenario is to enforce certain *inexpensive* controls (based on addressing and/or route) in real-time, (i.e., without expensive data authentication/integrity checks), combined with bulk packet-based accounting to later detect fraudulent (or unaccounted for) resource usage.

In summary, the design choices of object granularity and enforcement mode are more context-dependent than are enforcement location, protocol, or principal granularity.

3.6 Summary

We conclude this section with the results of the above discussion summarized in Table 2.[6].

Resources	Security services	Enforcement location	Enforcement protocol	Principal granularity
End-systems	All	End-systems, Access Control Servers	Transport and above	Any
Stub AD Network Resources	2,3,11,13	Border Routers, Access Control Servers	Network	End-system, AD, user class
Transit AD Network Resources	2,3,11,13	Border Routers, Access Control Servers	Routing	AD, user class
Route Selection	2,3	Border Routers, Route Servers	Routing and Route Computation	Any

Table 2: Policy Enforcement Parameters

[6]We have excluded the issues in Section 3.5 because the suggested approach is very context dependent and does not lend itself to accurate representation in a table.

4 Integrated View of Policy Enforcement

In Sections 2 and 3 we argued for dividing the discussion of security according to the type of resource being protected. However, from a technical and policy perspective, it is essential for these three levels of policy to be coordinated and consistent. To this end we present an integrated view of policy enforcement support. We will focus on three opportunities for integration – routers, specialized servers and encryption support.

4.1 Border Routers

As can be seen in Table 2, the heaviest burden of enforcement falls on the border routers. This is particularly true of border routers in stub ADs which need to implement controls at the granularity of both ADs and end-systems (as shown in Figure 4.1). For locally originated traffic, the border router needs to control: i) source end-system access to the outside, ii) destination AD and end-system, and iii) inter-AD route access. Traffic coming in from the outside has to be controlled on the same basis, with the exception of route selection (if routes are composed at the source). However, it is possible for a border router in a destination AD to refuse incoming traffic based on the route it followed.

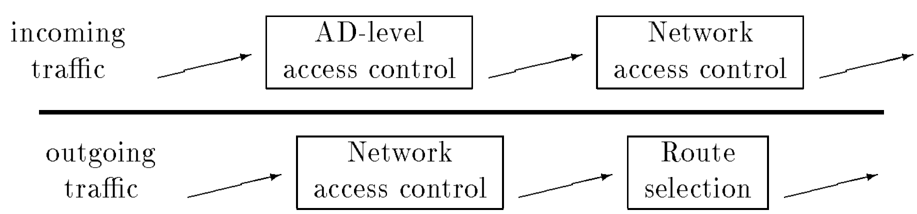

Figure 1: Border router controls for stub ADs.

4.2 Servers

Network managers need to create consistent policies across all three types of controls addressed. This may involve more than one server function. In a stub AD, network administrators will define which end-systems will be reachable from the outside. These policies are then enforced by the particular Access Control Server (ACS) used. The ACS controls the inclusion of particular principals and objects (e.g., end-system addresses) in router access control lists. For those end-systems that are made accessible, network administrators will want to monitor the end-system controls employed to make sure that unreachable end-systems are not vulnerable.

These same network administrators will create policies to be used in selecting routes. The same inter-AD relationships and trust levels that determine the end-system and network resource access policies will determine route selection criteria. Once routes are selected for possible use by the AD, route servers are responsible for handing out routes to end-systems (or border routers) within the AD. Therefore, network administrators must also specify access control lists to control distribution of the route servers' routes. A route server, therefore, is involved in all three types of policy – network resource control policies are considered in route computation, followed by route selection policies, and finally by end-system policies that control distribution of routes. Finally, for ADs that support some transit traffic as well as end-system access, transit routing policies, advertised via the inter-AD routing protocol, must be coordinated with stub-AD network resource controls.

In the case of transit ADs, there is less of a concern with coordinating several types of policy. However, network administrators will need tools to assist them in composing useful transit policies. Such policies must on the one hand realize the AD's control requirements, and on the other hand, be as compatible as possible with the policies of all other ADs. If no attention is paid to compatibility, then the transit service may be of little use to anyone in the internet. Support of policy routing is discussed further in [1, 7, 16].

4.3 Encryption support

Protocols that implement the security services for the enforcement of access control policies will have to make use of encryption to support authentication, data integrity, and confidentiality (if applicable). Therefore, there is some opportunity to take advantage of common support for encryption as well as key distribution and management.

For example, evolving standards for private electronic mail make use of a public key certificate model [2, 17]. This approach is well suited for the distribution of signed routing updates and to authenticated route setup. Key distribution will involve route servers and border routers and will be at a coarse granularity of ADs. Public Key certificates issued by a globally recognizable and trusted authority can be used to identify ADs that have no previous history of association. For example, a source AD can include its public key certificate as a part of route setup when one (or more) ADs in the route are being used for the first time.

4.4 Performance Impact

As is always the case, introduction of new controls influences communication. In particular, increased delay is due mostly to increased processing in border routers and increased packet length (used to carry authentication, data integrity and other related values). Previous results [6, 7] show that the impact due to the latter is greatly overshadowed by the additional processing in border routers, which is, in turn dominated by the cryptographic (signature) operations.

Recent development of fast non-cryptographic message integrity functions, e.g., MD4[21], has greatly reduced the cost of the integrity check computations. Authentication is attained by signing a short integrity value instead of the entire message. Furthermore, authentication and integrity methods based entirely on the use of one-way hash functions (i.e., completely avoiding cryptographic computation) are currently under development.[7]

[7] S. Kent, *Private Communication*, November 1990.

4.5 Conclusion

In conclusion, an integrated view of access control is needed in an environment of interconnected ADs in order to achieve effective and efficient placement of function for different types of policy, and to define consistent policies across the many network elements, protocols, and servers that are involved. In this brief paper, we proposed a framework for placement of access control functions. We applied the original *end-to-end* argument to network resources and concluded that resources other than sufficiently protected end-systems are best protected at the border routers. This argument is reinforced by the increasing concerns with respect to resource usage feedback and cost recovery that are raised by the commercialization of internetwork transit facilities [8].

References

[1] L. Breslau and D. Estrin, *Design of Inter-Administrative Domain Routing Protocols*, **Proceedings of 1990 ACM Sigcomm Conference**, September 1990.

[2] CCITT, *The Directory Authentication Framework*, **CCITT Recommendation X.509**, 1988.

[3] D. Clark, *Policy Routing in Internet Protocols*, **Journal of Internetworking: Research and Experience**, October 1990.

[4] D. Estrin, *Controls for Inter-Organization Networks*, **IEEE Transactions on Software Engineering**, February 1987.

[5] D. Estrin, *Policy Requirements for Inter Administrative Domain Routing*, **Computer Networks and ISDN Systems**, To appear in 1991.

[6] D. Estrin, J. Mogul, G. Tsudik, *Visa Protocols for Controlling Inter-Organizational Datagram Flow*, **IEEE Journal on Selected Areas in Communications**, May 1989.

[7] D. Estrin and G. Tsudik, *Secure Control of Transit Internetwork Traffic*, **Computer Networks and ISDN Systems**, To appear in 1991.

[8] D. Estrin and L. Zhang, *Design Considerations for Usage Accounting and Feedback in Internetworks*, **SIGCOMM Computer Communications Review**, October 1990.

[9] S. Hares, D. Katz *Administrative Domains and Routing Domains, a Model for Routing in the Internet*, **RFC 1136, SRI Network Information Center**, December 1989.

[10] IEEE 802.10, Standard for Interoperable LAN Security, **LAN Security Working Group, P802.10/D6**, September 1989.

[11] International Standards Organization, *Information Processing Systems – Open Systems Interconnection – Basic Reference Model*, **ISO 7498**, 1977

[12] International Standards Organization, *Security Architecture*, **ISO 7498-2-1988(E)**, 1988

[13] International Standards Organization, *OSI Routeing Framework*, **ISO/TF 9575**, 1989.

[14] P. Karger, *Authentication and Discretionary Control in Computer Networks*, **Journal of Computer Networks and ISDN Systems**, January 1986.

[15] S. Kent, *Protocols and Techniques for Data Communication Networks*, **Englewood Cliffs, NJ:Prentice Hall, pp. 369-432**, 1980.

[16] M. Lepp and M. Steenstrup, *An Architecture for Inter-Domain Policy Routing*, **BBN Report Number 7345**, July 1990, BBN Corporation, Cambridge, MA 02138.

[17] J. Linn, *Privacy Enhancement for Internet Electronic Mail. Part I: Message Encipherment and Authentication Procedures*, **RFC 1113, SRI Network Information Center**, January 1989.

[18] S. Mulender et al, *Amoeba: A Distributed Operating System for the 1990s*, **IEEE Computer**, Volume 23, No. 5, May 1990.

[19] D. Nessett, *Factors Affecting Distributed System Security*, **IEEE Transaction on Software Engineering, vol. SE-13**, 1987.

[20] J. Postel, *Internet Protocol*, **RFC 791, SRI Network Information Center**, September 1981.

[21] R. Rivest, *The MD4 Message Digest Algorithm*, **Proceedings of CRYPTO'90**, August 1990.

[22] R. Rivest, A. Shamir and L. Adleman, *A Method for Obtaining Digital Signatures and Public Key Cryptosystems*, **Communications of the ACM**, February 1978.

[23] J. Saltzer, D. Reed, and D. Clark, *End-to-End Arguments in System Design*, **ACM Transactions on Computer Systems**, 1984, pp. 195-206.

[24] J. Saltzer and M. Schroeder, *Protection of Information in Computer Systems*, **Proceedings of the IEEE**, September 1975.

[25] M. Satyanarayan, *Scalable, Secure, and Highly Available File Access*, **IEEE Computer**, Volume 23, No. 5, May 1990.

[26] M. Steenstrup and Open Routing Working Group, *Inter-Domain Policy Routing Specification*, **BBN Report Number 7346**, December 1990, BBN Corporation, Cambridge MA 02138.

[27] J. Steiner, *The Kerberos Network Authentication Service Overview*, **MIT Project Athena RFC, Draft 1**, April 1989.

[28] U.S. Department of Defense, *Trusted Computer System Evaluation Criteria*, **Standard 5200.28-STD**, December 1985.

[29] V. Voydock and S. Kent, *Security Mechanisms in High-level Network Protocols*, **ACM Computing Surveys**, June 1983.

[30] S. Walker, *Network Security: The Parts of the Sum*, **Proceedings of 1989 IEEE Symposium on Security and Privacy**, May 1989.