INTERNET DRAFT
Category: Informational
Title: draft-calhoun-diameter-framework-08.txt
Date: June 2000

Pat R. Calhoun
Sun Microsystems, Inc.
Glen Zorn
Cisco Systems, Inc.
Ping Pan
Bell Labs
Haseeb Akhtar
Nortel Networks

DIAMETER Framework Document

Status of this Memo

Abstract

Current Internet Service Providers (ISPs) scale their networks by using the RADIUS protocol, which provides user Authentication, Authorization and Accounting (AAA) of Dial-up PPP clients. The recent work done in the Roaming Operations (ROAMOPS) Working Group was to investigate whether RADIUS could be used in a roaming network, and concluded that RADIUS was ill-suited for inter-domain purposes.

The IETF has formed a new NAS Requirements Working Group, and part of their charter is to document the next generation NAS' AAA requirements. Recently, the Mobile-IP Working Group also documented their own AAA requirements that would help Mobile IP scale for Inter-Domain mobility.

The DIAMETER protocol is a follow-on to the RADIUS protocol. DIAMETER addresses the known RADIUS deficiencies, and is intended for use with the NASREQ, ROAMOPS and Mobile IP application space.

Table of Contents

1.0 Introduction

Historically, the RADIUS protocol has been used to provide AAA services for dial-up PPP [17] and terminal server access. Over time, routers and network access servers (NAS) have increased in complexity and density, making the RADIUS protocol increasingly unsuitable for use in such networks.

The Roaming Operations Working Group (ROAMOPS) has published a set of specifications [19, 20, 21] that define how a PPP user can gain access to the Internet without having to dial into his/her home service provider's modem pool. This is achieved by allowing service providers to cross-authenticate their users. Effectively, a user can dial into any service provider's point of presence (POP) that has a roaming agreement with his/her home Internet service provider (ISP), the benefit being that the user does not have to incur a long distance charge while traveling, which can sometimes be quite expensive.

Given the number of ISPs today, ROAMOPS realized that requiring each ISP to set up roaming agreements with all other ISPs did not scale. Therefore, the working group defined a "broker", which acts as an intermediate server, whose sole purpose is to set up these roaming agreements. A collection of ISPs and a broker is called a "roaming consortium". There are many such brokers in existence today; many also provide settlement services for member ISPs.

The Mobile-IP Working Group has recently changed its focus to inter administrative domain mobility, which is a requirement for cellular

carriers wishing to deploy IETF-based mobility protocols. The current cellular carriers requirements [22, 23] are very similar to the ROAMOPS model, with the exception that the access protocol is Mobile-IP [2] instead of PPP.

The DIAMETER protocol was not designed from the ground up. Instead, the basic RADIUS model was retained while fixing the flaws in the RADIUS protocol itself. DIAMETER does not share a common protocol data unit (PDU) with RADIUS, but does borrow sufficiently from the protocol to ease migration.

The basic concept behind DIAMETER is to provide a base protocol that can be extended in order to provide AAA services to new access technologies. Currently, the protocol only concerns itself with Internet access, both in the traditional PPP sense as well as taking into account the ROAMOPS model, and Mobile-IP.

Although DIAMETER could be used to solve a wider set of AAA problems, we are currently limiting the scope of the protocol in order to

ensure that the effort remains focussed on satisfying the requirements of network access. Note that a truly generic AAA protocol used by many applications might provide functionality not provided by DIAMETER. Therefore, it is imperative that the designers of new applications understand their requirements before using DIAMETER.

1.1 Requirements language

In this document, the key words "MAY", "MUST, "MUST NOT", "optional", "recommended", "SHOULD", and "SHOULD NOT", are to be interpreted as described in [9].

1.2 Terminology

Accounting
: The act of collecting information on resource usage for the purpose of trend analysis, auditing, billing, or cost allocation.

Authentication

The act of verifying the identity of an entity (subject).

Authorization
The act of determining whether a requesting entity (subject) will be allowed access to a resource (object).

AVP
The DIAMETER protocol consists of a header followed by objects. Each object is encapsulated in a header known as an Attribute-Value-Pair.

Broker
Although a DIAMETER proxy provides routing of DIAMETER authentication, authorization and accounting requests, a broker provides DIAMETER message routing while preserving strong security. A DIAMETER broker can also provide redirect services by providing a requesting DIAMETER server the information necessary to contact a target server directly.

DIAMETER Client
The DIAMETER Client is the device that users contact in order to get access to the network. An example of a client would be a Network Access Server (NAS) and a Foreign Agent (FA).

DIAMETER Server
The DIAMETER server is the device that provides for authentication

and authorization of a user or node requesting access to the network. The DIAMETER Server also collects accounting information for authenticated sessions.

Home Domain
This is the Internet service provider or corporate network with whom the user maintains an account relationship.

Integrity Check Value (ICV)
An Integrity Check Value is an unforgeable or secure hash of the message with a shared secret.

Interim accounting
An interim accounting message provides a snapshot of usage during a user's session. It is typically implemented in order to provide

for partial accounting of a user's session in the event of a device reboot or other network problem that prevents the reception of a session summary message or session record.

Session record
: A session record represents a summary of the resource consumption of a user over the entire session. Accounting gateways creating the session record may do so by processing interim accounting events or accounting events from several devices serving the same user.

Local Domain
: This is the Internet service provider whom the user uses in order to get access. Where roaming is implemented the local ISP may be different from the home ISP.

Network Access Identifier
: In order to provide for the routing of DIAMETER authentication and accounting requests, the userID field used in PPP and Mobile IP (known as the Network Access Identifier or NAI) and in the subsequent DIAMETER authentication and accounting requests, can contain structure. This structure provides a means by which the DIAMETER proxy will locate the DIAMETER server that is to receive the request. The NAI is defined in [3].

Proxy Server
: In order to provide for the routing of DIAMETER authentication and accounting requests, a DIAMETER proxy can be employed. To the NAS, the DIAMETER proxy appears to act as a DIAMETER server, and to the DIAMETER server, the proxy appears to act as a DIAMETER client.

Realm
: The portion of the Network Access Identifier [3] that immediately

follows the '@' character. As required in [3], NAI realm names are required to be unique, and are piggybacked on the administration of the DNS namespace. DIAMETER makes use of the realm, also loosely refered to as domain, to determine whether messages can be satisfied locally, or whether they must be proxied.

Real-time Accounting
: Real-time accounting involves the processing of information on

resource usage within a defined time window. Time constraints are typically imposed in order to limit financial risk.

Redirect Services
: A DIAMETER broker is said to provide redirect services by returning contact information to a requesting DIAMETER server in order to allow it to communicate directly with another server within a roaming consortium. The roaming consortium that allows for redirect services typically also provides certificate authority services in order to allow the end servers to communicate in a secure fashion.

Roaming relationships
: Roaming relationships include relationships between companies and ISPs, relationships among peer ISPs within a roaming association, and relationships between an ISP and a roaming consortia. Together, the set of relationships forming a path between a local ISP's authentication proxy and the home authentication server is known as the roaming relationship path.

Session
: The DIAMETER protocol is session based. When an authentication request is initially transmitted, it includes a session identifier that is used for the duration of the session. The Session-Identifier AVP contains the identifier and must be globally unique.

2.0 Problems to be addressed

The RADIUS protocol was designed in the early 1990's as an attempt to solve a scaling problem associated with dial-in and telnet servers. Over time the networks became more complex (e.g. roaming networks) and the Network Access Servers (NAS) increased in complexity and density. These changes combined with a massive deployment of the protocol uncovered some fundamental issues with the protocol that needed to be fixed. The DIAMETER protocol was designed as a next generation RADIUS protocol, designed with roaming and high density NASes in mind.

This section will describe the documented, and undocumented, RADIUS

problems known today. Further sections will describe how the DIAMETER protocol addresses each one of these problems.

2.1 Strict limitation of attribute data

One of problems that RADIUS suffers from is its inherent limitation on the length of attribute data. This limitation is imposed by the fact that the protocol's attribute header only reserves one byte for the length field. The RADIUS protocol does specify that larger data can be spanned across multiple attributes, however doing so introduces a new set of problems. The RADIUS protocol also allows multiple attributes of the same type to be included within a message. Therefore, it is difficult for a RADIUS server, or client, to determine whether multiple identical attributes are in fact multiple independent attributes, or a single fragmented attribute.

2.2 No documented retransmission procedure

The RADIUS protocol states that the identifier field, found within the header, is used to identify retransmissions. This one byte field imposes a strict limitation on the number of requests that can be pending at any given time to 255. In the early 1990's, this number was sufficient, but the increased density of most NASes today make the protocol nearly unusable. Most NASes today have fixed this problem by including information in other attributes to bypass this limitation. However, the RADIUS servers have also had to support this change in protocol since they must be able to properly identify retransmissions. The RADIUS protocol also states that the identifier MUST be changed if any data is changed in a request.

For this reason, most RADIUS servers keep a cache of received RADIUS request (e.g. all messages received in the last 60 seconds). The RADIUS servers then attempt to match some attributes within the received requests with all attributes in all messages in the cache. This places a very heavy burden on the RADIUS servers, but unfortunately is the only method of identifying retransmissions given the fact that the RADIUS protocol does not have any good scheme. This hack has proved necessary since some NASes have had to change some information within requests in the retransmission queue (such as session length).

2.3 Inability to control flow to servers

Given the rather bursty nature of the RADIUS protocol, current

servers have no way of properly managing their receive buffers. This is in part due to the fact that RADIUS operates over UDP, and does not include any windowing support. This has been known to cause large bursts of requests to be directed to a server, which can burden a server's ability to respond in a timely manner. This problem is most prevalent in cases where a server becomes unavailable and all requests must be sent to an alternate server, or when an ingress port on the NAS becomes available (e.g. T3 port on NAS).

2.4 End to end message acknowledgment

The RADIUS protocol requires that a NAS retransmit a request until a successful or failed response is received, and does not permit a RADIUS server to retransmit a response. This is problematic since there are many times when a server does receive a request, but cannot respond before the NAS determines that the request must be retransmitted. This can occur for many reasons, including the fact that processing a RADIUS request, which includes authentication and authorization of the user, a database lookup and logging events, can be lengthy.

2.5 No retransmission procedure

Another reason why NASes typically retransmit is when a SERVER receives a large number of requests, and cannot process all of them in a timely manner. The side effect here is that if the NAS retransmits requests to the server, it simply causes further damage to the busy server. Since the RADIUS server cannot retransmit, some RADIUS servers keep a cache of responses sent in the past 60 seconds in order to minimize processing should a retransmission be received. As previously discussed, identifying a retransmission is a very CPU intensive tasks, but perhaps not quite as intensive as a database lookup.

2.6 Heavy processing cost

The introduction of proxy RADIUS network have made this acknowledgement scheme even worse, since the end server must respond in a timely manner. Each intermediate RADIUS server adds additional latency to proxied requests due to the application processing cost. This has been known to cause unnecessary retransmission of requests by NASes, which impose heavy burden on the proxies, and the network.

When a NAS retransmits a request a maximum number of times, it assumes that the server is no longer available and transmits the

message to an alternate server. If there are many messages in the retransmission queue, all other requests are also transmitted to the new server. Since a burst of requests were sent to the server, the chances that it can satisfy all requests before the retransmission period are very small, which causes unnecessary retransmissions.

2.7 Silent discarding of packets

The RADIUS protocol states that messages that do not contain the expected information, or messages that have errors are silently discarded. Silently discarding messages can create a serious problem since no response is sent to the NAS, which then has to assume that the server is no longer reachable. Since proxy networks are transparent to a NAS, should a server in a proxy chain silently discard a request, it will cause the NAS to assume that the local (first hop) server is no longer available.

2.8 No fail-over server support

Most NASes today support a large number of RADIUS servers in an attempt to provide resilience. However, the RADIUS protocol itself makes this very difficult due to the problems described above. Since a NAS does not know a priori whether a specific server is available, when it switches to an alternate server, it must retransmit a message a maximum number of times before determining that the server in question is down, and that the next server in the configuration chain must be tried. Taking an example of a NAS with 8 servers configured, if the next 3 servers in the configuration chain were down, it would take the NAS x number of seconds to reach an available server (where x is equal to the retransmission interval * the maximum number of retransmissions * 3), which is most often longer than the clients are willing to wait.

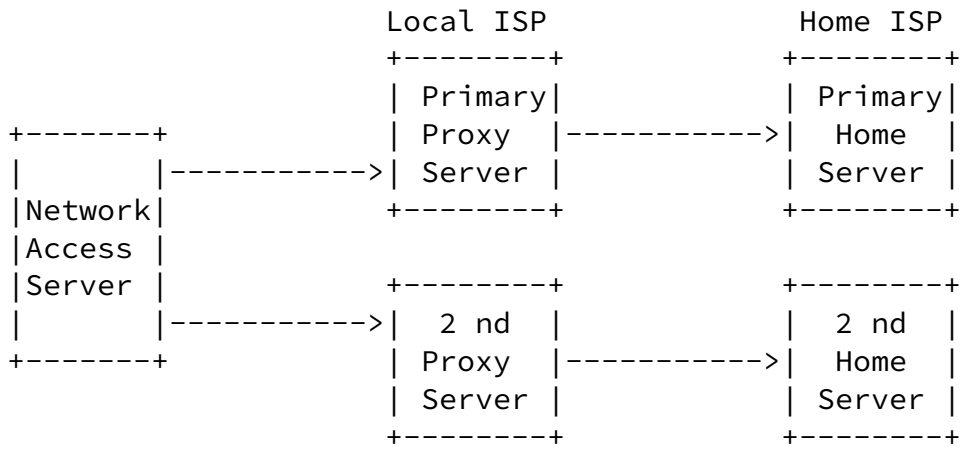

Figure 1: RADIUS Proxy Network

2.9 client/server protocol

Given that a RADIUS server cannot know a priori whether a downstream RADIUS server is reachable, and the fact that the NAS must retransmit any messages, the RADIUS protocol is not well suited to proxy environments. Since servers are not aware of a peer's reachability, most RADIUS networks are designed by creating parallel links between primary and alternate servers (see figure 1). In this example the local ISP's primary server communicates with the home ISP's primary server, while the 2nd servers communicate directly. When the NAS issues a request to the primary server, the first hop server attempts to proxy the request to the primary server at the home network. The NAS will attempt to retransmit the request n number of times, and the primary server will simply forward the request to the primary server at the home network.

Should no response be received, the primary server could attempt to forward the request to the 2nd server at the home network, but since the NAS is controlling the retransmissions, it may not have the opportunity to do so. Therefore, the NAS will redirect all requests to the local ISP's 2nd server. Given the protocol's limitations, it requires a large number of RADIUS servers in order to provide resilient service.

The above problem is further aggravated should the local ISP attempt to proxy to two different administrative network's servers. Take an example where the local ISP issues two authentication requests, one for abc.net and another for xyz.com. Let's also assume that abc.net's primary server is down, while xyz's 2nd server is down. Should such a problem occur, all requests for abc.net would cause the NAS to switch to the local ISP's 2nd server, while all requests to xyz.net would cause the NAS to switch back to the local ISP's primary server. This

clearly illustrates that the RADIUS protocol cannot be reliably used in proxy networks.

2.10 No unsolicited messages

The RADIUS protocol does not allow a server to send unsolicited messages to the NAS. As network services became more complex, this limitation has forced manufacturers to break the RADIUS protocol in this area in order to allow servers to communicate with the client. This is widely used for accounting purposes, and to allow a server to inform a NAS that a session should be terminated. Unfortunately, the lack of a standard method of doing this has caused many non-interoperable implementations to be deployed.

2.11 Authentication Replay Attacks

In today's PPP world, the NAS provides a challenge to the user, which is then computed by the PPP user to create the challenge response. This is commonly known as CHAP [26], and is a popular PPP authentication scheme. Before roaming networks existed, service providers would own both the NAS and the RADIUS server and this wasn't considered a problem. However, now that the NAS and the RADIUS server are owned by two separate administrative domains, the fact

that the non-trusted NAS generates a challenge provides the ability for authentication replay attacks. A NAS, or any RADIUS server in a proxy chain, can have access to a valid challenge/response pair, which can be replayed at a later time.

The EAP protocol [10], which will be supported as part of RADIUS extensions can solve this problem, but the fact that EAP is not widely deployed on clients, and that many EAP authentication transforms cannot be used within RADIUS (due to the limitation on attribute data size) makes it difficult to use. Furthermore, given the RADIUS protocol's requirement for end-to-end retransmissions, since some EAP authentication types involve a higher number of round trips than what RADIUS currently supports, RADIUS and EAP cannot be used on networks that exhibit data loss. This is primarily due to the fact that most EAP (PPP) clients timeout before the authentication can be completed.

2.12 Hop-by-Hop security

The RADIUS protocol uses hop-by-hop security, which means that every hop in a RADIUS proxy network adds authentication data that is used by the next peer in the chain. RADIUS has no facility for strong

security, where security is maintained from the requestor and the responder, even though a request is handled by intermediate nodes. This has caused opportunities for fraud in RADIUS networks, since intermediate nodes can easily modify information (e.g. accounting information), and such events are untraceable.

2.13 No support for vendor-specific commands

Although the RADIUS protocol does support vendor-specific attributes, it does not allow for vendor-specific commands. This has caused serious inter-operability problems since vendors simply choose command identifiers at random, which can collide with other manufacturer's implementation.

2.14 No alignment requirements

Unlike most newer IETF protocols, the RADIUS protocol does not impose any alignment requirements, which adds an unnecessary burden on most processors. All fields within the header and attributes must be treated as byte aligned characters.

3.0 DIAMETER Architecture

The DIAMETER architecture consists of a base protocol and a set of protocol extensions (such as strong security, NASREQ, Mobile-IP and accounting). Functionality common to all supported services is implemented in the base protocol, while application-specific functionality may be provided through the extension mechanism.

The base protocol [18] must be supported for all DIAMETER applications, and defines the basic PDU format, a few primitives and the basic security services offered by the protocol. Unlike RADIUS, the DIAMETER protocol operates over SCTP [28], which provides reliability and an agressive retransmission and timeout mechanism. Additionally, DIAMETER defines a fail-over strategy, which is lacking in the RADIUS protocol. SCTP provides a windowing scheme, which allows the AAA servers to limit the flow of incoming packets. This can then be used by the AAA clients to distribute the traffic load across multiple servers. The transport layer's aggressive retransmission and timeout timers allow clients and servers to detect the reachability state of peers, allowing for quick transition to back-up servers.

As previously discussed, the ROAMOPS model introduces the AAA broker, which acts as an intermediate server redirecting requests to user's

home ISPs. ROAMOPS also described a set of attacks that one could mount if such a network was built using the RADIUS protocol [21]. In order to provide secure broker services, strong security is required at the application layer, since messages traverse application gateways (brokers).

The DIAMETER Strong Security Extension defines a set of extensions to the base protocol that provide authentication, confidentiality and non-repudiation at the Attribute-Value-Pair (AVP) level. With these extensions, it is possible to secure portions of a DIAMETER message, while other parts of the message are not secured. Secured objects are

called protected AVPs; non-secured objects are called unprotected AVPs. Using DIAMETER, brokers can add, delete or modify unprotected AVPs in a message.

The RADIUS protocol provides dial-up PPP AAA services by providing three commands and many Attributes. Attributes in RADIUS are analogous to AVPs in DIAMETER. In order to ease migration from RADIUS to DIAMETER, the first 256 AVPs in the DIAMETER AVP space are reserved for RADIUS compatibility. This allows both protocols to share a common dictionary and policy rules for PPP user profiles. Recently, the RADIUS protocol adopted support for the Extensible Authentication Protocol (EAP) [10], but RADIUS' lack of support for large attributes and its inherent unreliability has made the integration of the protocols very difficult.

The DIAMETER NASREQ Extension defines a set of authentication/authorization commands, which can be used for CHAP, PAP and EAP. DIAMETER's support for larger AVPs and the SCTP transport properties have made the use of EAP much more palatable, allowing for end-to-end user authentication, which reduces many of authentication replay attacks currently documented.

Unlike PPP, Mobile-IP hosts do not have a long-lived "nailed-up" connection to a PPP server, but rather get service from routers that provide service in a particular cell. In the Mobile-IP world, the router is known as a Foreign Agent, while the moving hosts are known as Mobile Nodes. The mobile node's home network has a host that forwards all messages destined to the mobile node through the Foreign Agent. This router is commonly referred to as the Home Agent.

Mobile-IP [7] allows the mobile nodes to move from one cell (subnet) to another while retaining the same IP address, minimizing the impact to applications. Although the Mobile-IP protocol could be deployed in a small network with any AAA services, a larger network suffers from many scaling issues such as:

- Static mobile node home address

- Static mobile node home agent
- Requirement to pre-configure mobile node profile on home agents
- No inter-domain mobility

Both PPP and Mobile-IP require that usage data be collected for uses such as capacity planning and for accounting purposes. The current standard protocol for accounting is SNMP [12], but experience indicates that SNMP often is not the correct protocol for service accounting. Today many applications and services use RADIUS accounting [4] as their accounting protocol, however the RADIUS accounting protocol is not an IETF standard; in addition, it suffers from similar scaling and security problems. The DIAMETER accounting extension [11] is designed to allow accounting information to be sent across administrative domains (optionally through brokers), and has been derived from an accounting requirements document [6, 8].

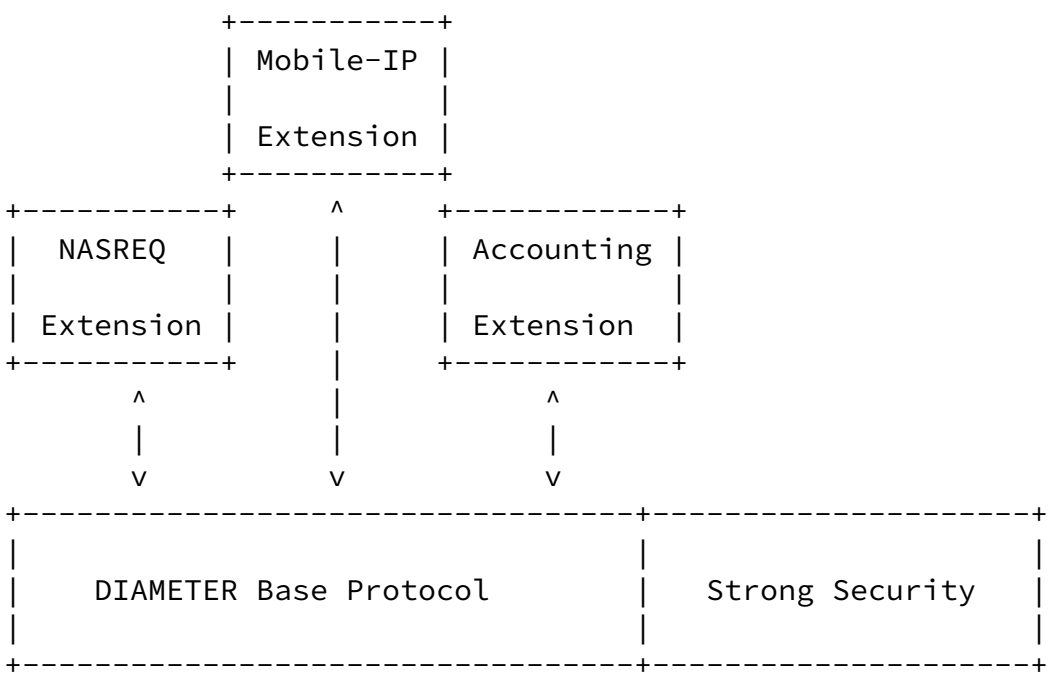

Figure 2: DIAMETER Protocol Architecture

3.1 DIAMETER Base Protocol

The Base Protocol defines the DIAMETER message format, a set of primitives and how the messages are transmitted in a secure fashion. The Base Protocol assumes a peer-to-peer communication model, as opposed to a client-server model. The following goals motivated the design of the base protocol:

- lightweight and simple to implement protocol
- Large AVP space
- Efficient encoding of attributes, similar to RADIUS

- Support for vendor specific AVPs and Commands
- Support for large number of simultaneous pending requests
- Reliability provided by underlying SCTP
- Well-defined fail-over scheme
- Ability to quickly detect unreachable peers
- No silent message discards
- Support of unsolicited messages to "clients"
- integrity and confidentiality at the AVP level
- Hop-by-Hop security
- One session per authentication/authorization flow
- Provide redirect (referal) services, to allow bypassing of broker

The DIAMETER base protocol is intended to simply provide a secure transport for the messages defined in the various application-specific extensions. It is therefore imperative that the base be lightweight and simple to implement.

In the DIAMETER protocol, data objects are encapsulated within the Attribute Value Pair (AVP). An AVP consists of three parts: the Identifier, Length and Data. A unique AVP Identifier is assigned to all data objects in order to be able to distinguish the data contained. The AVP Identifier namespace must be sufficiently large to ensure that future protocol extensibility is not limited by the size of the namespace, as in the RADIUS protocol. Furthermore, vendors wishing to add "proprietary" extensions must be allowed to do so by using a vendor-specific namespace, managed by IANA.

For many years the question as to whether RADIUS should operate over UDP or TCP has led to heated discussion. It must be determined whether the benefits that UDP provides are worth the implementation complexities. Over time, it has become clear that these benefits are well worth the cost. The issue with TCP is that an AAA protocol requires a quick retransmission and fail-over scheme, which TCP cannot provide. The DIAMETER protocol must be able to operate over a transport that has an aggressive retransmission strategy in order to efficiently switch to an alternate host when the peer in question is no longer reachable.

Contrary to RADIUS, the DIAMETER protocol requires that each node in a proxy chain acknowledge a request, or response, at the "transport" layer. Since DIAMETER operates over SCTP, which provides a reliable transport, each node in a proxy chain is responsible for retransmission of unacknowledged messages.

The SCTP transport provides retransmission detection, which greatly simplifies server implementations, and consequently allows a given

server to support a much larger number of transactions per second.

SCTP also provides windowing, which allows the flow of packets to a specific server to be controlled. Clever implementations can then decide to send the packets to an alternate server that can handle the load.

With the exception of a few security related errors, the DIAMETER protocol requires that all messages be acknowledged, either with a successful response or one that contains an error code.

Where the RADIUS protocol is client-server, the DIAMETER protocol is peer to peer, allowing unsolicited messages to be sent to NASes. There are many benefits to peer-to-peer AAA protocols. One example is the on-demand retrieval of accounting data; another, server-initiated session termination.

The Base DIAMETER protocol provides for hop-by-hop security, similar to the scheme employed by RADIUS today. However, the DIAMETER protocol also provides for replay protection through a timestamp mechanism. This security scheme requires a long lived security association to be established by peers, or can make use of keying material negotiated out of band. The Base Protocol also allows the built-in security measure to be turned off, (i.e., in cases where IPSec is in use).

The DIAMETER protocol is a session-oriented protocol, meaning that for each user being authenticated, there exists a session between the initiator of the authentication/authorization request and the home DIAMETER server. Sessions are identified through a session identifier, which is globally unique at any given time. All subsequent DIAMETER transactions (e.g. accounting) must include the session identifier to reference the session. A Session termination message exists in order to end a DIAMETER session, and all sessions have a timeout value in order to ensure that they can be cleaned up properly.

Since today's processors work more efficiently when objects are aligned on a 32-bit boundary, the DIAMETER protocol requires 32-bit alignment of all headers and the data. This has recently become a common requirement for many new protocols at the IETF.

3.1.1 Proxy Support

The DIAMETER protocol was designed from the beginning to support roaming networks. This means that every node in the network is responsible for it's own retransmissions, and the protocol does allow each node to know a priori the reachability state of each peer. This allows for a resilient network, and efficient retransmission scheme.

Figure 3 depicts a network where each DIAMETER server can communicate with all other servers.

Figure 3 depicts an example of a DIAMETER network that includes two proxy servers in the local network for resilience. Once a message has been sent from the NAS to one of its local proxy servers, they are responsible for any retransmissions of the message to one of the home servers. Since the underlying transport provides quick peer failure detection, upon such notification, the local proxies can quickly transmit the message to the alternate peer in the home network.

Figure 3 depicts an example of a proxy network that includes alternate servers for resilience. Each node in the proxy chain is responsible for its own retransmissions and fail-over detection. This provides the following benefits:

- The number of DIAMETER nodes in the network is greatly reduced
- The latency involved in switch-over to an alternate peer is greatly reduced
- Reliability is increased

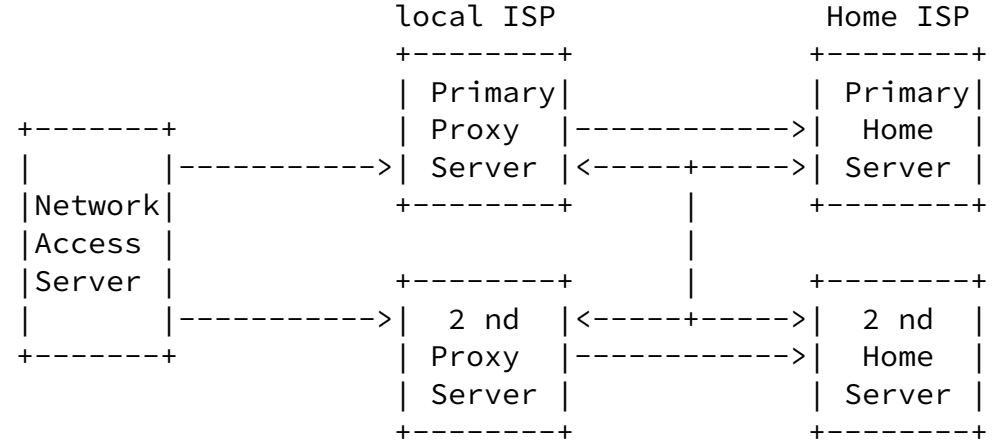

Figure 3: DIAMETER Proxy Network

3.1.2 Broker Support

A broker is a proxy server that provides simple DIAMETER message "routing" functions. Brokers are generally deployed in order to reduce the configuration information that would otherwise be necessary on all servers owned by ISPs within a roaming consortium. Brokers can provide two separate functions depending upon the business model.

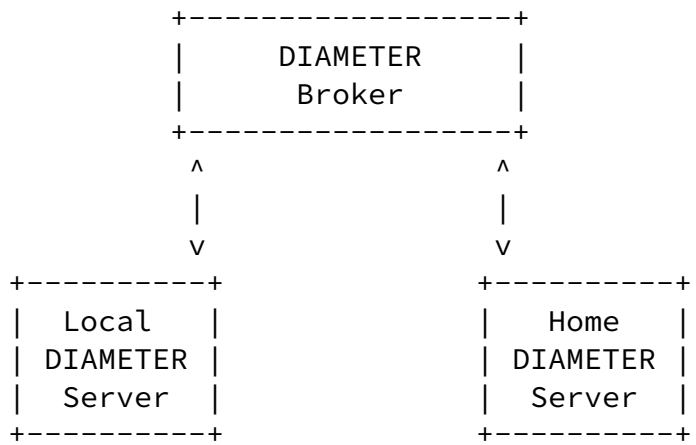

Figure 4: DIAMETER Roaming Consortium

The first where the broker forwards messages to the proper destination based on the NAI information (figure 4). In such a network, when the broker receives a request from a DIAMETER server, it determines the message's destination and can optionally perform some authorization decisions based on local policy.

The DIAMETER broker's organization can also provide Certificate Authority services, by issuing certificates to all DIAMETER servers within the consortium, or use existing certificates owned by DIAMETER servers. This allows the broker and the DIAMETER servers to

communicate in a secure fashion, without the need for long-lived secrets. Protocols such as IP Security [8] can allow for short-lived session keys to be generated and periodically refreshed.

The second broker model allows the end DIAMETER servers to directly communicate (figure 5). In this model the broker simply provides redirect services, which is aimed at reducing the amount of configuration that would otherwise be necessary on all end DIAMETER servers. When a DIAMETER servers sends a request the broker, the broker returns contact information that is then used by the requesting server to re-issue the request directly to the home DIAMETER server. In order for the end DIAMETER servers to be able to communicate in a secure fashion, a pre-established security association is required. This can be in the form of a long-lived shared secret, which has scaling problems, or via certificates when the broker's organization provides CA services. In the event that the broker also provides settlement services, it is possible for the accounting information, signed by both parties, to be transmitted to the broker by the server providing service to the user.

When the broker provides the message forwarding functions, it can validate that the source and destination DIAMETER servers are in

"good standing", which reduces the processing on the end servers. This can be done by having the broker check the server's certificates against a CRL, via an online certificate status protocol [25], or through local configuration. The broker can optionally attach the source server's certificate if it isn't already present in the message. When a broker receives a request from or destined to a realm that is either unrecognized or no longer part of the roaming consortium, an error will be returned to the requesting server.

The very fact that the DIAMETER servers in the roaming network do not have to burden themselves with validating certificates against a CRL, or some other certificate validation infrastructure, is a huge advantage. In cases of inter-consortium roaming, the brokers involved can be responsible for validating any certificates involved. Note that it is also possible for the broker to periodically issue new certificates to the roaming consortium members out-of-band in order to eliminate the need to add certificates to each message, decreasing the message size and the per-message processing penalty.

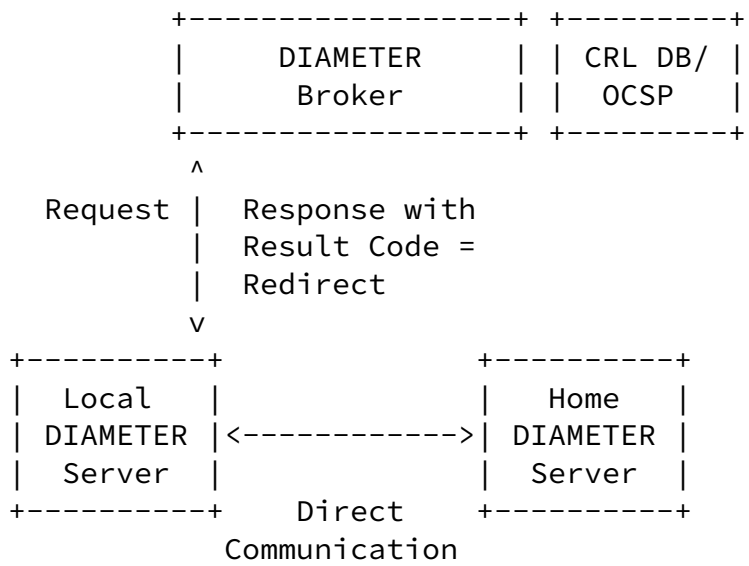

Figure 5: DIAMETER Broker Returning Redirect Indication

When the broker provides redirect services, the broker can return both the source and the destination server's certificates. The certificates are encapsulated within a DIAMETER attribute, and include a timestamp, an expiration time all signed by the broker. Should the end server's policy be setup such that they will trust the certificate returned by the broker, they do not have to make any additional certificate validation checks. However, local policy may require that the end DIAMETER servers validate periodically.

Note that even though some broker's do allow direct communication, some will require that all accounting messages be forwarded by the

broker. This is typically required when the broker also provides settlement services. In such a network, the broker normally requires some reassurances that the user was in fact authenticated and authorized by the home DIAMETER server prior to accepting accounting records. This can be achieved by requiring that both DIAMETER servers sign the Accounting data in a serial fashion [27].

3.2 Strong Security Extension

The DIAMETER base protocol allows DIAMETER servers to communicate securely, using hop-by-hop authentication. Hop-by-hop authentication

means that the requesting server has secure communication with the broker, and the broker has secure communicate with the destination server.

The Strong Security extension [27] provides strong authentication of selective AVPs, which MAY be used for repudiation purposes. This extension also allows for secure communication through intermediate DIAMETER proxies.

The extension achieves this functionality by allowing the Cryptographic Message Syntax (CMS) S/MIME object to be encapsulated within a DIAMETER AVP. The CMS object MAY be used for authentication, confidentiality and to carry certificates and certificate revocation lists (CRLs). The extension also provides for multi-party signatures, which is useful in environments where two or more parties must sign information, such as an accounting record.

DIAMETER clients, such as NASes and routers, aren't expected to implement strong security. This specification is targeted for the first hop proxy servers, and this functionality is normally only required when requests must traverse administrative domain boundaries.

The strong security extension MUST only be used in networks that include a Public Key Infrastructure (PKI).

3.3 Mobile-IP Extension

The Mobile-IP protocol is used to manage mobility of an IP host across IP subnets [7]. Recent activity within the Mobile-IP Working Group has defined the interaction between Mobile-IP and AAA in order to provide:

- Better scaling of security associations
- Mobility across administrative domain boundaries

- Dynamic home agent assignment

The Mobile IP protocol [7] works well when all mobile nodes belong to the same administrative domain. Some of the current work within the Mobile IP Working Group is to allow Mobile IP to scale across

administrative domains. This work requires modifications to the existing Mobile IP trust model.

Figure 6 depicts the DIAMETER trust model for Mobile-IP. In this model each network contains mobile nodes (MN) and a DIAMETER server. Each mobility device shares a security association (SA) with the DIAMETER server within its own home network. This means that none of the mobility devices initially share a security association. The DIAMETER servers in both administrative domains can either share a direct security association, or can have a security association with an intermediate broker.

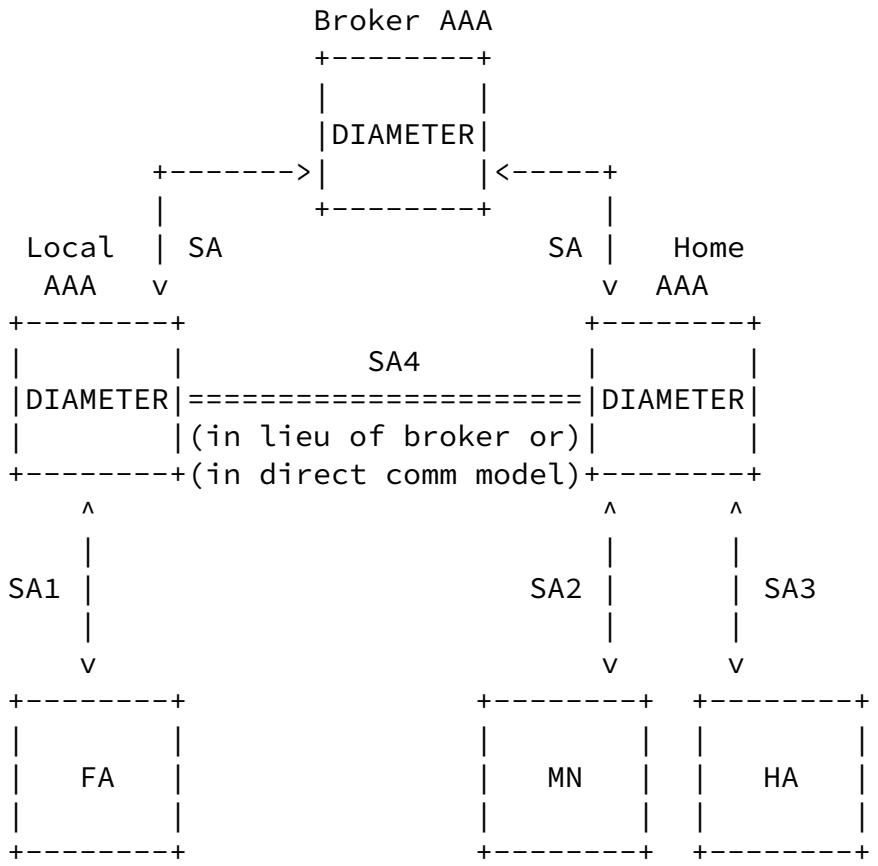

Figure 6 - Mobile-IP AAA Trust Model

Figure 7 provides an example of a Mobile-IP network that includes DIAMETER. In the integrated Mobile-IP/DIAMETER Network, it is assumed that each mobility agent shares a security association between itself and its local DIAMETER server. Further, the Home and Local DIAMETER servers both share a security association with the broker's DIAMETER server. Lastly, it is assumed that each mobile node shares a trust

relationship with its home DIAMETER Server.

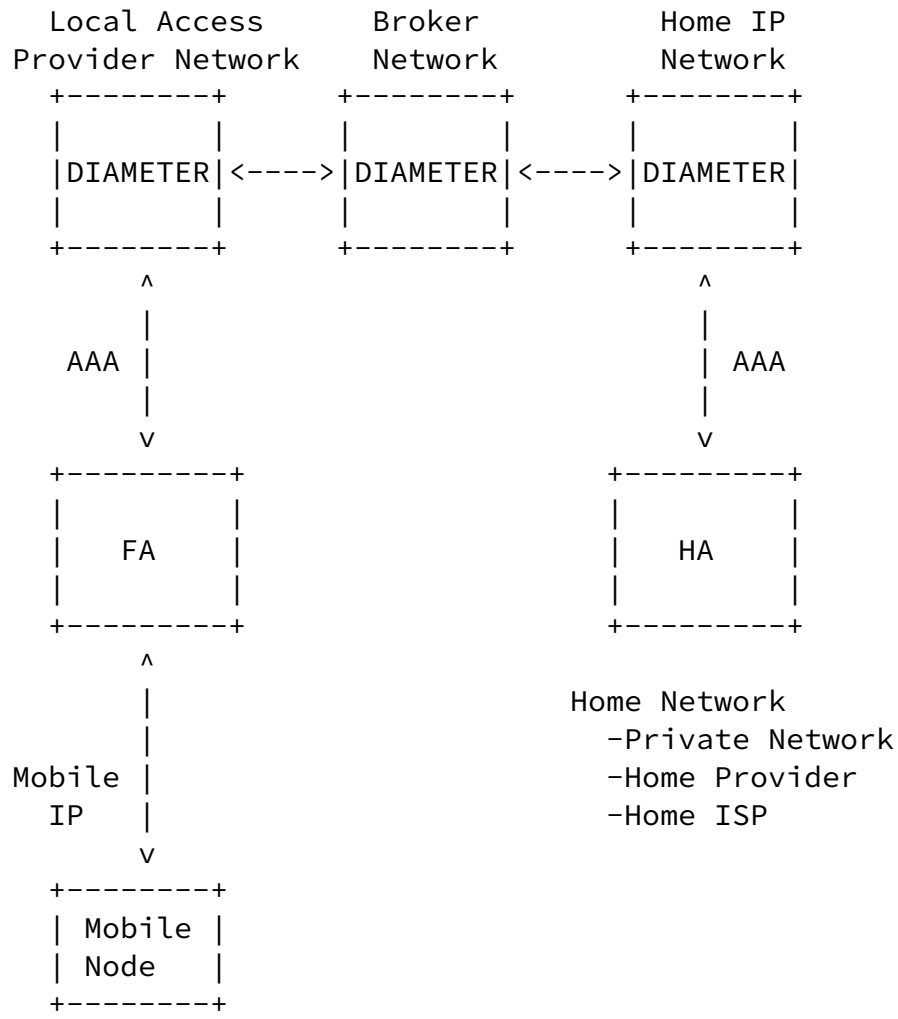

Figure 7 - General Wireless IP Architecture for Mobile-IP AAA

In this example, a Mobile Node appears within a local network and issues a registration to the Foreign Agent. Since the Foreign Agent does not share any security association with the Home Agent, it sends a DIAMETER request to its local DIAMETER server, which includes the authentication information and the Mobile-IP registration request. The Mobile Node cannot communicate directly with the home DIAMETER Server for two reasons:

- It does not have access to the network. The registration request is sent by the Mobile Node to request access to the network.
- The Mobile Node may not have an IP address, and may be requesting that one be assigned to it by its home provider.

The Local DIAMETER Server will determine whether the request can be satisfied locally through the use of the Network Access Identifier

[3] provided by the Mobile Node. The NAI has the form of user@realm and the DIAMETER Server uses the realm portion of the NAI to identify the Mobile Node's home DIAMETER Server. If the Local DIAMETER Server does not share any security association with the Mobile Node's home DIAMETER Server, it may forward the request to its broker. If the broker has a relationship with the home network, it can forward the request, otherwise a failure indication is sent back to the Local DIAMETER Server.

When the home DIAMETER Server receives the DIAMETER Request, it authenticates the user and begins the authorization phase. The authorization phase includes the generation of:

- Dynamic session keys to be distributed among all mobility agents
- Optional dynamic assignment of a home agent
- Optional dynamic assignment of a home address (note this could be done by the home agent).
- Optional assignment of QOS parameters for the mobile node [22]

Once authorization is complete, the home DIAMETER Server issues an unsolicited DIAMETER request to the Home Agent, which includes the information in the original DIAMETER request as well as the authorization information generated by the home DIAMETER server. The Home Agent retrieves the Registration Request from the DIAMETER request and processes it, then generates a Registration Reply that is sent back to the home DIAMETER server in a DIAMETER response. The message is forwarded through the broker back to the Local DIAMETER server, and finally to the Foreign Agent.

The DIAMETER servers maintain session state information based on the authorization information. If a Mobile Node moves to another Foreign Agent within the local administrative domain, a request to the local DIAMETER server can be done in order to immediately return the keys that were issued to the previous Foreign Agent. This eliminates an additional round trip through the internet when micro mobility is involved, and enables smooth hand-off. In order for the DIAMETER server to be able to provide the keying information to the new Foreign Agent, they must have a pre-existing security association.

Note that smooth hand-off is really a mobility function, and it is not clear that DIAMETER should be involved. However, this example is provided for completeness.

If the Mobile Node enters a service area owned by a new service provider, the authentication and authorization request will have to be sent back to the home DIAMETER server, which will create new keying information.

3.3.1. Minimized Internet Traversal

Although it would have been possible for the DIAMETER interactions to be performed for basic authentication and authorization, and the Registration flow to be sent directly to the Home Agent from the Foreign Agent, one of the key Mobile-IP DIAMETER requirements is to minimize Internet traversals. Including the Registration Request and Replies in the DIAMETER messages allows for a single traversal to authenticate the user, perform authorization and process the Registration Request. This streamlined approach is required in order to minimize the latency involved in getting wireless (cellular) devices access to the network. New registrations should not increase the connect time more than what the current cellular networks provide.

3.3.2. Key Distribution

In order to allow the scaling of wireless data access across administrative domains, it is necessary to minimize the security associations required. This means that each Foreign Agent does not share a security association with each Home Agent on the Internet. The Mobility Agents share a security association with their local DIAMETER server, which in turn shares a security association with other DIAMETER servers. Again, the use of brokers (as defined by ROAMOPS) allows such services to scale by allowing the number of relationships established by the providers to be reduced.

After a Mobile Node is authenticated, the authorization phase includes the generation of Sessions Keys. Specifically, three keys are generated:

- K1 Key to be shared between the Mobile Node and the Home Agent
- K2 Key to be shared between the Mobile Node and the Foreign Agent

- K3 Key to be shared between the Foreign Agent and the Home Agent

Each key is encrypted in two separate methods. K1 is encrypted using SA3 (for the Home Agent), and using SA2 (for the Mobile Node). K2 is encrypted using SA4 (for the Foreign Agent) and using SA2 (for the Mobile Node). Lastly, K3 is encrypted using SA4 (for the Foreign Agent), and using SA3 (for the Home Agent). When the Foreign DIAMETER Server receives the keys, they are decrypted and re-encrypted using SA1. All of the Security Associations (SAx) are shown in figure 6. The keys destined for the foreign and home agent are propagated to the mobility nodes via the DIAMETER protocol, while the keys destined for the Mobile Node are sent via the Mobile-IP protocol.

Figure 8 depicts the new security associations used for Mobile-IP message integrity using the keys derived by the DIAMETER server.

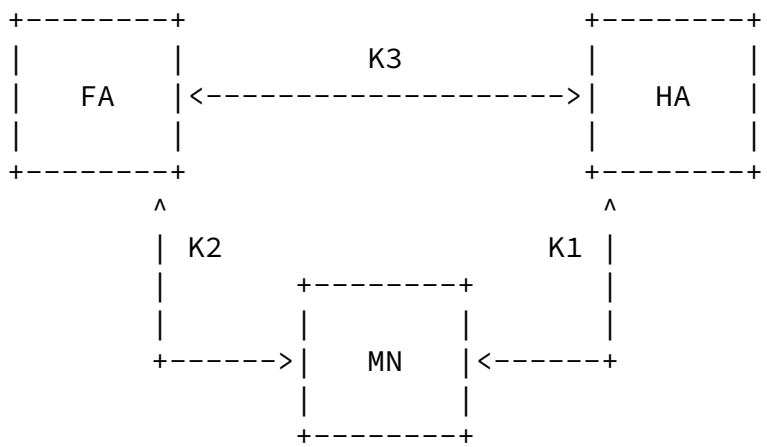

Figure 8 - Security Association after Key Distribution

Once the session keys have been established and propagated, the mobility devices can exchange registration information directly without the need of the DIAMETER infrastructure. However the session keys have a lifetime, after which the DIAMETER infrastructure must be used in order to acquire new session keys.

3.4 NASREQ Extension

The NASREQ extension provides authentication and authorization for dial-in PPP users, terminal server access and tunneling applications,

such as L2TP. The extension makes use of the attributes defined in the RADIUS protocol to carry the data objects. This was intended to ease migration of existing RADIUS servers to DIAMETER since they could share a single dictionary and user profile. Furthermore, this would reduce the amount of processing required for an inter-working system that acts as a RADIUS/DIAMETER bridge.

DIAMETER has native EAP support that solves known problems in the RADIUS protocol. Furthermore, DIAMETER takes end-to-end authentication one step further by providing for end-to-end authentication via PPP's CHAP. This allows for a more secure authentication infrastructure without having to replace or modify the installed base of clients.

If end-to-end CHAP is used in bridged DIAMETER/RADIUS environments, the bridge host is responsible for generating the challenge to the user.

The remaining authentication and authorization logic found in RADIUS implementations can then be re-used. The basic changes are the

message formats and the transmission mechanism as defined in the DIAMETER base protocol. This section does not detail RADIUS authentication and authorization. The interested reader should refer to RFC 2138.

3.5 Accounting Extension

The Accounting extension provides usage collection to both the Mobile-IP and the NASREQ extensions. The accounting requirements specifications [6, 8] define that an accounting protocol must provide the following functionality:

- Negotiable transfer mechanism.
- Provide general purpose AVPs.
- Flexible to allows new extensions to use the accounting extension.
- Scalable to allows millions to users and thousands of sites.
- Secure accounting data transfer.

The DIAMETER protocol encodes the actual accounting information using

the Accounting Data Interchange Format (ADIF) [24]. ADIF was intended to allow a uniform encoding of accounting data to be transferred over virtually any transport (e.g. DIAMETER, SMTP, HTTP, etc).

The DIAMETER Accounting Extension allows accounting information to be sent in two modes; real-time and batched. Real-time accounting transfers are useful in environments where timely arrival of the information is required, such as when debit cards are used. Batched accounting transfers are useful in environments that do not need up to the minute accounting records. However, it is possible that in inter-domain networks, real-time accounting data delivery will be more popular since the ISPs involved will want to receive some guarantees of payment prior to providing service.

The DIAMETER protocol is session oriented, and each session typically has a finite lifetime. Prior to the timeout of a session, a user typically needs to be re-authentication and/or re-authorized in order to extend the life of the session. In the Mobile-IP world, this equates to the mobility registration lifetime, while in PPP this means that the PPP authentication must be re-opened. When a re-authentication and/or re-authorization occurs, a new token is generated, which is used in the corresponding accounting message.

The DIAMETER Accounting extension combined with the Strong Security [27] extension (see section 3.2), provides strong authentication of accounting data, which MAY be used for repudiation purposes. The strong security extension also allows multiple parties to sign the

accounting information, which is beneficial in environments that include a referral broker. The foreign and home servers can both sequentially sign the accounting record, and submit the result to the broker. The broker can then use the signatures to ensure that both parties agreed to the contents of the accounting record.

3.6 Resource Management

Many network access services requiring AAA support have a requirement for servers that maintain session state information. An example of such a requirement is in the dial-up PPP world. With the introduction of flat-rate internet access, there has been a surge in fraud where a user provides his username/password pair to other people. The end

result is that a single username (account) can have simultaneous concurrent sessions.

Internet Service Providers have had to implement proprietary extensions to protocol, such as RADIUS, in order to attempt to identify when such fraud occurs. Unfortunately, since the protocol does not provide the necessary functionality required to maintain state information, these solutions have been unreliable.

The DIAMETER Base Protocol [18], the Accounting extension [11], the Mobile IP [13] and NASREQ [23] extensions provide some of the functionality that is required for servers to maintain state information, such as:

- Reliable Transport
- Indication of the termination of a session
- A Reboot message
- Interim Accounting
- Accounting On/Off message
- Ability to re-authorize an existing session

Although the above features do allow nodes to maintain state information, it is necessary for a DIAMETER node to request a snapshot of active sessions from a peer. This may be used when state information is lost, which could occur after a device failure, or this may be done periodically in order to ensure that the state is current.

The DIAMETER Resource Management extension [5] provides the messages that are required for a node to request a snapshot of active sessions from a peer. State information is exchange via the Resource-Token AVP, which is used to encapsulate a set of AVPs that describe the session and resources used. There is one Resource-Token AVP for each active session.

3.7 DIAMETER Command Naming Conventions

The following conventions are proposed for the naming of Diameter messages. Diameter commands typically start with an object name, and end with one of the following verbs:

3.7.1 Request/Answer

Request is used when the command is asking the peer to do something for it, for example, set up a session, or reconfigure some parameters. The Answer MUST contain either a positive or negative result code, telling the requester whether or not the request successfully occurred. Other information can also be returned in the Answer.

For example, AA-Request asks the peer device to authorize and/or authenticate a user in order to set up a session. The request may fail, thus the answer may be positive or negative.

3.7.2 Query/Response

Query is used when the command is asking for information that it expects the peer to have. An example would be querying for current configuration information, or querying for information on resources or sessions in use. The Response usually contains a positive result code and the information, or a negative result code with the reason for not answering the query.

For example, Resource-Query requests the peer device to return specific information about one or more resources. The answer is returned in a Resource-Response.

3.7.3 Indication

Indication is used when the command is giving information on something that is about to or has already occurred. The peer receiving the message does not respond to the message, but a transport level acknowledgement must be done in order to ensure that the message was reliably delivered.

For example the base draft defines a message that is used to ensure that a peer is still active. The Device-Watchdog-Ind message has no associated response, but is acknowledged by the underlying transport.

4.0 Why not LDAP?

One common question is whether LDAP would provide the functionality required.

A Server MAY wish to access policies using LDAP, but the use of LDAP between the client and the server is not possible. The use of LDAP in this case would require that all routers have read/write access to the directory. Most customers would not accept this requirements and it is not efficient.

In the case of roaming, customers would have to open up their directory so outside routers have writable access. The security implications set aside, having 1000's of routers constantly read/write to the directory would cause some additional problems to the Directory Service.

Finally, LDAP does not provide server initiated messages which is a requirement for an AAA protocol.

5.0 References

[1] Rigney, et alia, "RADIUS", RFC-2138, Livingston, April 1997

[2] Veizades, Guttman, Perkins, Kaplan, "Service Location Protocol", RFC-2165, June 1997.

[3] Aboba, Beadles, "The Network Access Identifier", RFC 2486, January 1999.

[4] Rigney, "RADIUS Accounting", RFC-2139, April 1997.

[5] P. Calhoun, N. Greene, "DIAMETER Resource Management", draft-calhoun-diameter-res-mgmt-03.txt, IETF Work in Progress, April 2000.

[6] B. Aboba, J. Arkko, D. Harrington. "Introduction to Accounting Management", draft-ietf-aaa-acct-05.txt, IETF work in progress, June 2000.

[7] C. Perkins, Editor. IP Mobility Support. RFC 2002, October 1996.

[8] S. Kent, R. Atkinson, "Security Architecture for the Internet Protocol", RFC 1825, November 1998.

[9] Bradner, "Key words for use in RFCs to Indicate Requirements Levels", BCP 14, RFC 2119, March 1997.

[10] L. Blunk, J. Vollbrecht, "Extensible Authentication Protocol (EAP)", RFC 2284, March 1998.

[11] J. Arkko, P. Calhoun, P. Patel, G. Zorn, "DIAMETER Accounting Extension", draft-calhoun-diameter-accounting-06.txt, IETF work in progress, June 2000.

[12] J. Case, D. Harrington, R. Presuhn, B. Wijnen, "Message Processing and Dispatching for the Simple Network Management Protocol:", RFC 2572, April 1999.

[13] P. Calhoun, C. Perkins, "DIAMETER Mobile IP Extensions", draft-calhoun-diameter-mobileip-08.txt, IETF work in progress, June 2000.

[14] M. Baum, H. Perritt, "Electronic Contracting, Publishing and EDI Law", Prentice-Hall, ISBN 0-471-53135-9.

[15] P. Calhoun, C. Perkins "Mobile IP Foreign Agent Challenge/Response Extension", draft-ietf-mobileip-challenge-12.txt, IETF work in progress, June 2000.

[16] D. Harkins, D. Carrell, "The Internet Key Exchange (IKE)" RFC 1409, November 1998.

[17] W. Simpson, "The Point-to-Point Protocol (PPP)", RFC 1661, STD 51, July 1994.

[18] P. Calhoun, A. Rubens, H. Akhtar, E. Guttman, "DIAMETER Base Protocol", draft-calhoun-diameter-15.txt, IETF work in progress, June 2000.

[19] B. Aboba, G. Zorn, "Criteria for Evaluating Roaming Protocols", RFC 2477, January 1999.

[20] B. Aboba, J. Lu, J. Alsop, J. Ding, W. Wang, "Review of Roaming Implementations", RFC 2194, September 1997.

[21] B. Aboba, J. Vollbrecht, "Proxy Chaining and Policy Implementation in Roaming", RFC 2607, June 1999.

[22] T. Hiller and al, "CDMA2000 Wireless Data Requirements for AAA", draft-hiller-cdma2000-aaa-01.txt, IETF work in progress, June 2000.

[23] P. Calhoun, W. Bulley, A. Rubens, J. Haag, "DIAMETER NASREQ Extension", draft-calhoun-diameter-nasreq-03.txt, IETF work in progress, April 2000.

[24] B. Aboba, D. Lidyard, "The Accounting Data Interchange Format (ADIF)", draft-roamops-acctng-07.txt, IETF work in progress, August 1999.

[25] Myers, Ankney, Malpani, Galperin, Adams, "X.509 Internet Public Key Infrastructure Online Certificate Status Protocol (OCSP)", RFC 2560, June 1999.

[26] W. Simpson, "PPP Challenge Handshake Authentication Protocol (CHAP)", RFC 1994, August 1996.

[27] P. Calhoun, W. Bulley, S. Farrell, "DIAMETER Strong Security Extension", draft-calhoun-diameter-strong-crypto-03.txt, IETF work in progress, April 2000.

[28] R. Stewart et al., "Simple Control Transmission Protocol", draft-ietf-sigtran-sctp-09.txt, IETF Work in Progress, April 2000.

6.0 Acknowledgements

The Authors would like to thanks Bernard Aboba and Jari Arkko for their Accounting Requirements contribution. Thanks also goes to Erik Guttman for some very useful comments in helping make this draft more readable. The Mobile-IP Extension section was text originally written by Pat Calhoun for another Internet-Draft, which was subsequently cleaned up by Dave Spence. The authors would like to thank Nenad Trifunovic, Tony Johansson and Pankaj Patel for their participation in the Document Reading Party. A final thanks to Stephen Farrell for his security review.

7.0 Author's Addresses

Questions about this memo can be directed to:

Pat R. Calhoun
Sun Laboratories, Network and Security
Sun Microsystems, Inc.
15 Network Circle
Menlo Park, California, 94025
USA

Phone: +1 650-786-7733
Fax: +1 650-786-6445
E-mail: pcalhoun@eng.sun.com

Glen Zorn
Cisco Systems, Inc.
500 108th Avenue N.E., Suite 500
Bellevue, WA 98004
USA

Phone: +1 425 438 8218
E-Mail: gwz@cisco.com

Ping Pan
Bell Laboratories
Lucent Technologies

101 Crawfords Corner Road
Holmdel, NJ 07733
USA

Phone: +1 732-332-6744
E-mail: pingpan@dnrc.bell-labs.com

Haseeb Akhtar
Wireless Technology Labs
Nortel Networks
2221 Lakeside Blvd.
Richardson, TX 75082-4399
USA

Phone: +1 972-684-8850
E-Mail: haseeb@nortelnetworks.com

8.0 Full Copyright Statement

RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.