INSTITUTE FOR DEFENSE ANALYSES

Wide Area Network Acceleration in a High Assurance Enterprise

William R. Simpson

Kevin E. Foltz

July 3, 2015

IDA Non-Standard
NS D-5404
Log: H15-000020
Copy

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

The Institute for Defense Analyses is a non-profit corporation that operates three federally funded research and development centers to provide objective analyses of national security issues, particularly those requiring scientific and technical expertise, and conduct related research on other national challenges.

About This Publication

This work was conducted by the Institute for Defense Analyses (IDA) under contract HQ0034-14-D-0001, Task BC-5-2283, "Architecture, Design of Services for Air Force Wide Distributed Systems," for USAF HQ USAF SAF/CIO A6. The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

Wide Area Network Acceleration in a High Assurance Enterprise

Kevin Foltz and William R Simpson

***Abstract*— Bandwidth continues to be a problem for the active enterprise. One solution to bandwidth problems over long-haul distances with restricted bandwidth is the Wide Area Network (WAN) Accelerator. This accelerator works by tokenizing blocks of information that are sent multiple times in network traffic. Because many such communications include previously transmitted material, the accelerator traffic quickly damps out to transmissions that include tokens instead of the original communication. The tokens are reconstituted before delivery, and the receiver has a seamless connection and is unaware of the process. The acceleration is not without its drawbacks. The process does not work on encrypted traffic due to the random nature of encryption. For high assurance systems using an end-to-end paradigm, there are two main areas of concern. The first is security (how do we handle the decryption/re-encryption process?). The second is integrity (how do we maintain end-to-end integrity when encryption is broken?)**

This paper discusses the current approach to WAN acceleration and the changes that are required by a high assurance end-to-end approach. The latter rely on a well-formed security paradigm for the enterprise.

***Index Terms* — Network Acceleration Appliance, Protection, IT Security, Encryption, Key Management, Wide Area Network, Integrity**

I. Introduction

A WAN accelerator [1] is an appliance that optimizes bandwidth to improve the end user's experience on a wide area network (WAN). The appliance, which can be a physical hardware component, a software program, or an appliance running in a virtualized environment, speeds up the time it takes for information to flow back and forth across the WAN by using compression and data deduplication techniques to reduce the amount of data that needs to be transmitted. An accelerator works by caching duplicate files or parts of files so they can be referenced instead of having to be sent across the WAN again. Many of the products have evolved beyond the core acceleration techniques. The WAN optimization controllers (WOC) further optimize the WAN link by accounting for known problems with common network protocols. Protocol optimization cleans up chatty protocols used in common enterprise standards such as Common Internet File System (CIFS), Microsoft Exchange, and even TCP/IP to eliminate the typical overhead found in these communication protocols. These optimizations require a deeper understanding of the protocols and can be accomplished only through significant collaboration with application vendors or reverse engineering by the WAN accelerator vendor.

WAN optimization encompasses [2]:

- Traffic shaping, in which traffic is prioritized and bandwidth is allotted accordingly.
- Data deduplication, which reduces the data that must be sent across a WAN for remote backups, replication, and disaster recovery.
- Compression, which shrinks the size of data to limit bandwidth use.
- Data caching, in which frequently used data are hosted locally or on a local server for faster access.
- Monitoring the network to detect non-essential traffic.
- Creating and enforcing rules about downloads and Internet use.
- Protocol spoofing, which is a method of bundling chatty protocols so they are, in effect, a single protocol.

WAN optimization vendors include Blue Coat Systems, Cisco, Expand Networks, F5 Networks, Juniper, and Riverbed Technology.

II. Current WAN Accelerator Approaches

Acceleration of unencrypted traffic is not of interest in the high assurance end-to-end paradigm, so we will concentrate on the steps involved in handling encrypted traffic. Initially, there is a need to decrypt the traffic since tokenization is ineffective for encrypted traffic. This is due to the randomization of bit streams that are a property of encryption. This is normally done by passing the private key of the server to the server-side WAN accelerator appliance, as shown in Fig. 1.

As shown in Fig. 2, the Transport Layer Security (TLS) keys are computed and then passed between the accelerator units. The premaster secret keying material that is passed is protected by the private key of the server. The sharing of the server's private key allows the server-side appliance to extract the premaster secret and compute the session's master secret, just as the endpoints compute it. This is used to generate the encryption keys and Message Authentication Code (MAC) secrets for each TLS connection within this session. As shown in Fig. 1, the session keys and MAC secrets are then transmitted to the client-side WAN accelerator.

Manuscript received January 1, 2015; revised January 21, 2015. This work was supported in part by the U.S. Secretary of the Air Force and The Institute for Defense Analyses (IDA). The publication of this paper does not indicate endorsement by any organization in the Department of Defense or IDA, nor should the contents be construed as reflecting the official position of these organizations.

K. Foltz is with the Institute for Defense Analyses. (email: kfoltz@ida.org).

William R. Simpson is with the Institute for Defense Analyses, 4850 Mark Center Drive, Alexandria, Virginia 22311 USA and is the corresponding author phone: 703-845-6637, FAX: 703-845-6848 (e-mail: rsimpson@ida.org).

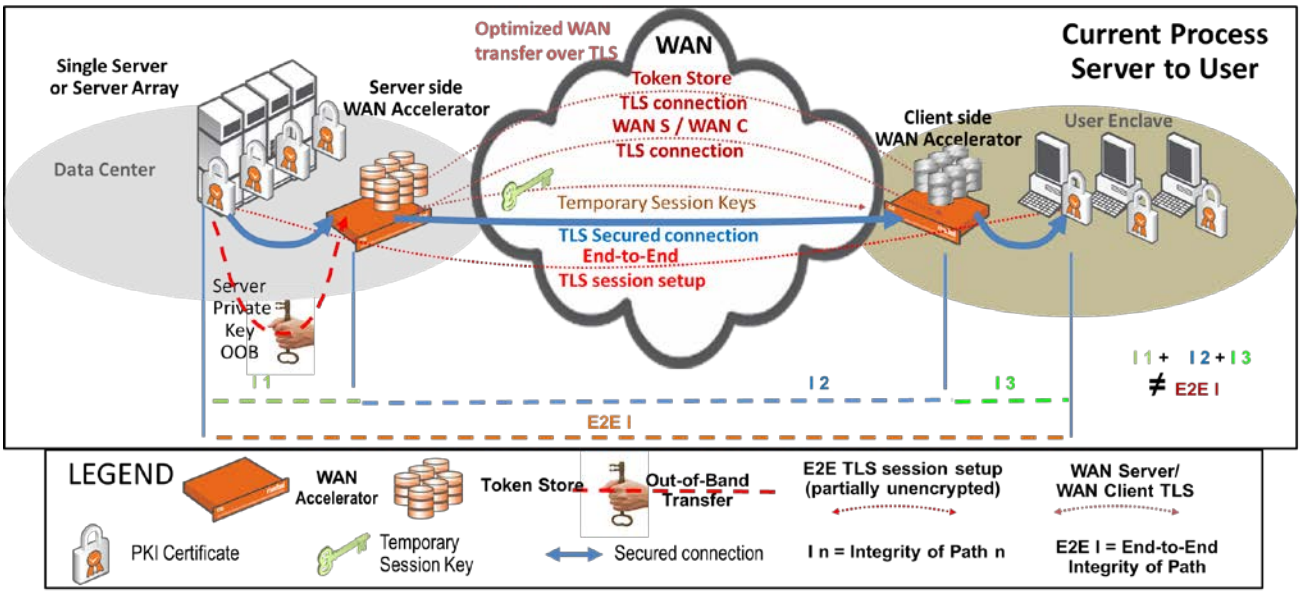

Fig. 1. WAN Accelerator Current Process Server to User

A MAC is a cryptographic checksum on data that use a shared secret key to detect both accidental and intentional modifications of the data (for integrity). With this MAC secret and the encryption key and Initialization Vectors (IV), the client-side WAN accelerator can send TLS content that can be decrypted and validated properly by the client.

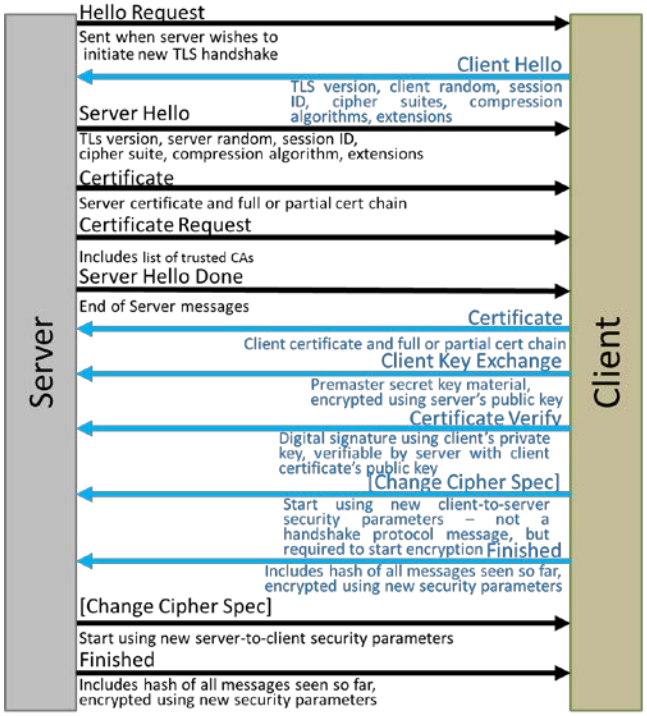

Fig. 2. TLS Handshake

While sharing of private keys is an easy way to provide the appliances with visibility to the content, it is a singularly bad idea from a security standpoint. First, in high assurance systems, the private keys are locked in a Hardware Storage Module (HSM) and are not shareable. Second, loss of a private key will compromise identity, break integrity and confidentiality of all TLS traffic to the server (past, present, and future), allow an adversary to impersonate the entity on an existing connection, and allow the adversary to impersonate the entity on new connections initiated by the adversary. In contrast, loss of a *session* key will entail a loss of only *confidentiality* for a *single connection* within a TLS session, with no risks for identity impersonation.

Fig. 1 shows the basic flows. This includes the logical end-to-end client-to-server TLS connection and its three secure component connections to, between, and from the WAN accelerators. It also includes TLS sessions to provide the session keys to the client-side accelerator, to transmit the tokenized packets to the client-side accelerator, and optionally to keep the tokenization stores synchronized. Although shown as separate TLS connections between the same endpoints, these may instead be implemented as separate logical flows within a single TLS connection. Aside from key management issues, the difference is not important.

Fig. 1 also shows the concern of integrity. The piecewise integrity of the component connections (I1, I2, and I3) does not amount to overall end-to-end integrity (E2E I).

$$\mathbf{I\,1 + I\,2 + I\,3 \neq E2E\ I} \qquad \mathbf{(1)}$$

In an end-to-end security paradigm, the goal is for the receiver to verify that the message sent from the server is identical to the message received at the client. Even when TLS MACs are included in all component connections, the integrity of the end-to-end transmission is not preserved unless we undertake some extraordinary measures. When a component connection retransmits data with a new MAC value, any changes made by that component to the original data will go undetected. In order to avoid explicit trust of the WAN processes, the TLS reconstruction of the server traffic by the client-side WAN accelerator must be modified. To understand this, we must examine the way in which TLS messages are formed.

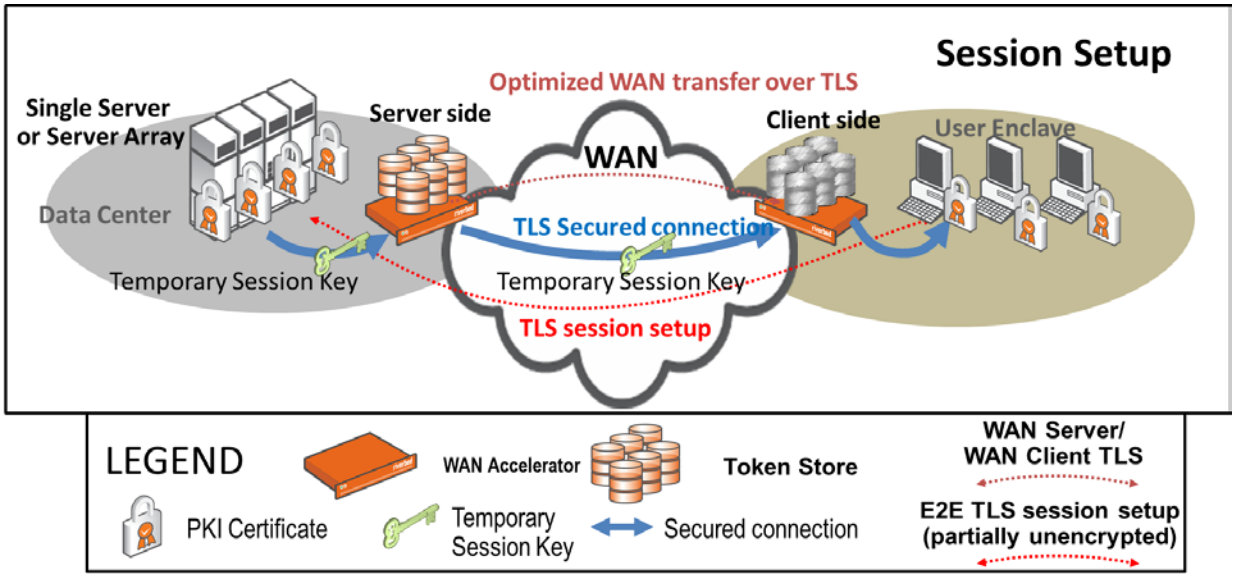

Fig 3. Alternative Encrypted Web Server Communications

III. An Alternative to Private Key Passing

For most interactions using enterprise level security approaches, traffic content does not need to be inspected. Firewall functionality will still be available using the headers that are not encrypted. However, certain functions, including WAN acceleration, require content inspection. For these conditions we recommend an alternative to sharing private keys as follows:

1. Web application shares only the TLS session keys that are needed for each appliance to function
2. No shared private keys – each active entity has its own unique public/private key pair.
3. Web application is endpoint for browser requests.

Fig. 3 shows the alternative recommendation. HTTP traffic is encrypted using TLS from browser to web application. The WAN Accelerator uses the provided session keys to decrypt TLS traffic. The balance of the transaction works in the same way, with the exception of the integrity problem. For this we must examine the TLS package development process.

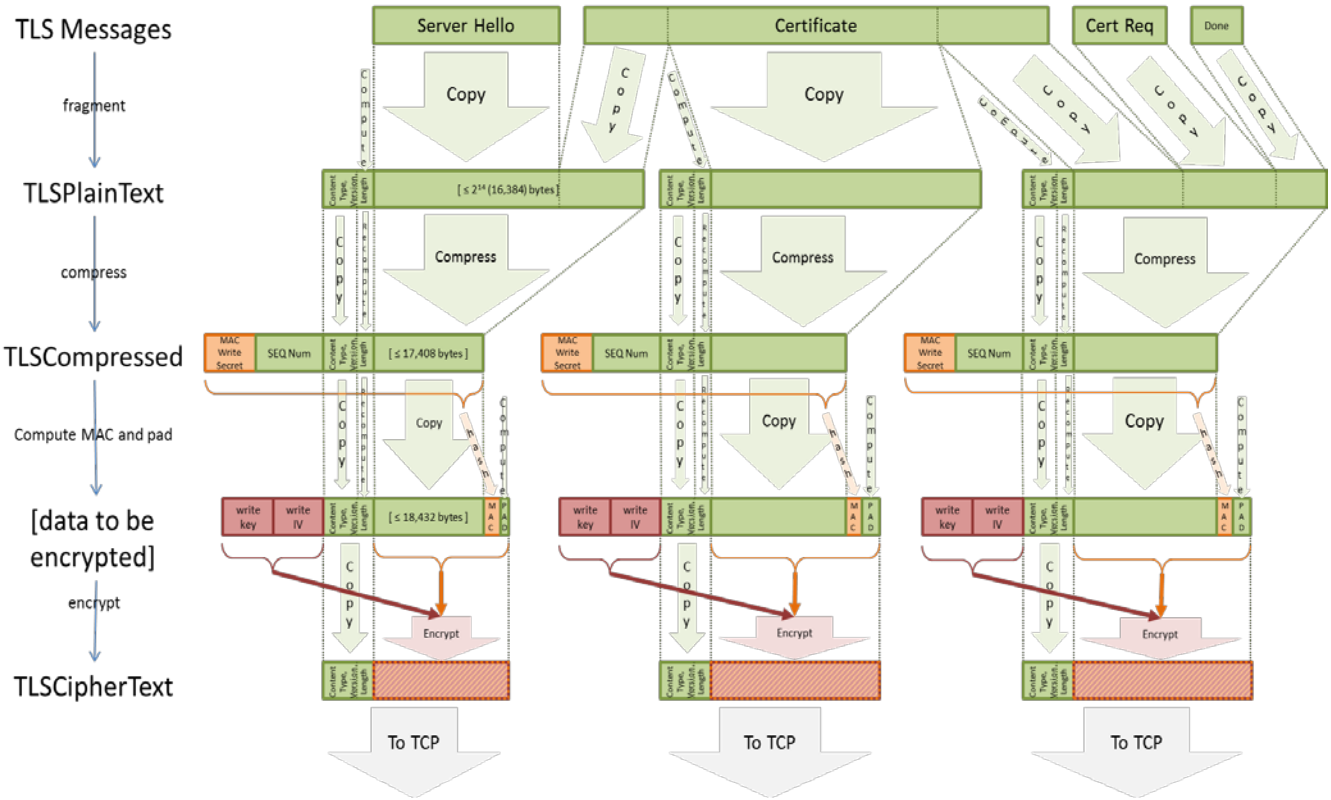

Fig. 4. TLS Transmission Steps

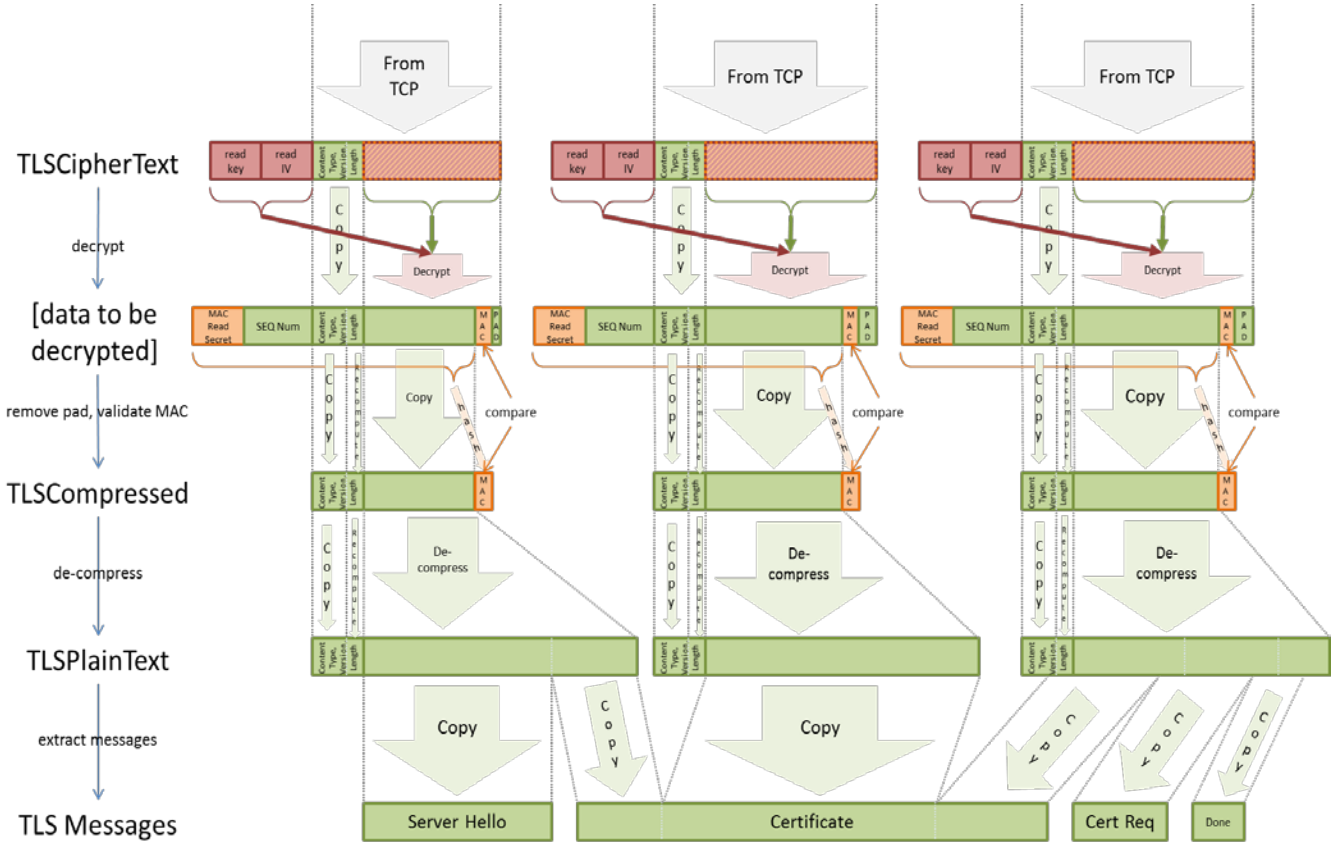

Fig. 5. TLS Reception Steps

IV. Integrity in a TLS Session

The message breakdown and TLS packet construction [19-21] are shown in Fig. 4, including fragmentation of messages, addition of headers, compression, addition of the MAC and padding, and encryption. The MAC secret and the values that depend on it are indicated in orange. The encryption keys and IVs, and the values that depend on them, are indicated in red. The ciphertext has both red and orange, since it relies on both the MAC secret and the encryption keys and IVs.

Fig. 5 shows the reverse process, where the receiver decrypts, validates the received MAC, decompresses, extracts content, and defragments into the original messages.

One important feature of TLS is that the encryption key/IV and MAC secret are separate values with separate functions. With only the encryption key and IV, an intermediate node can view the content but not modify it. It cannot compute a valid MAC for the modified content without the MAC secret. Also relevant is that the fragmentation boundaries of the original messages are important for MAC computation. Identical messages that are fragmented differently will have different MACs since the MACs depend on the fragments to which they are appended.

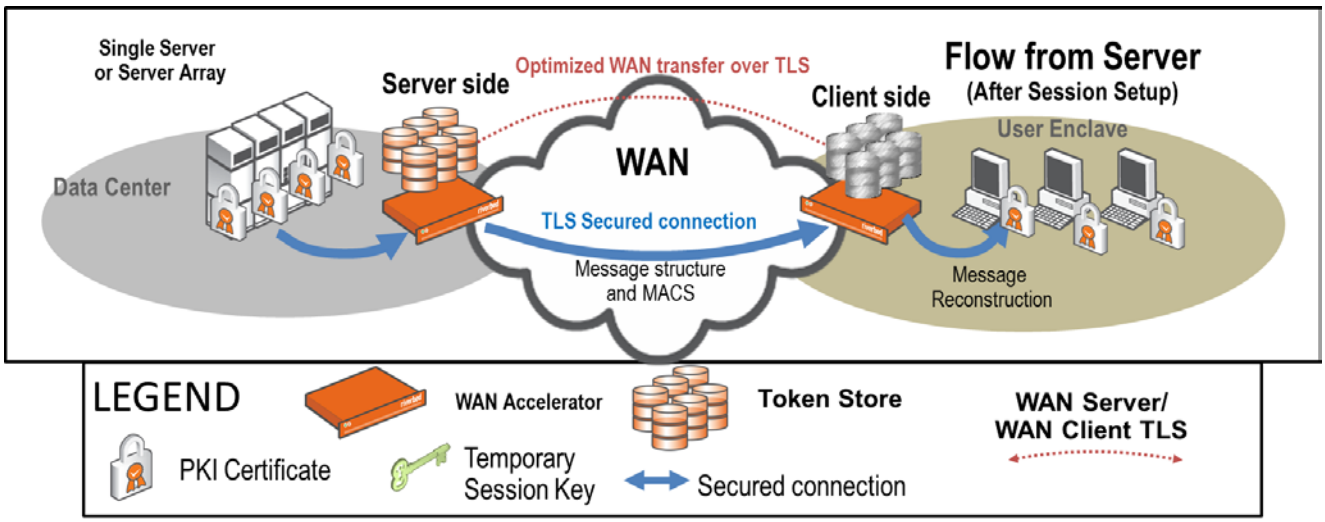

Fig. 6. Flow from Server Side

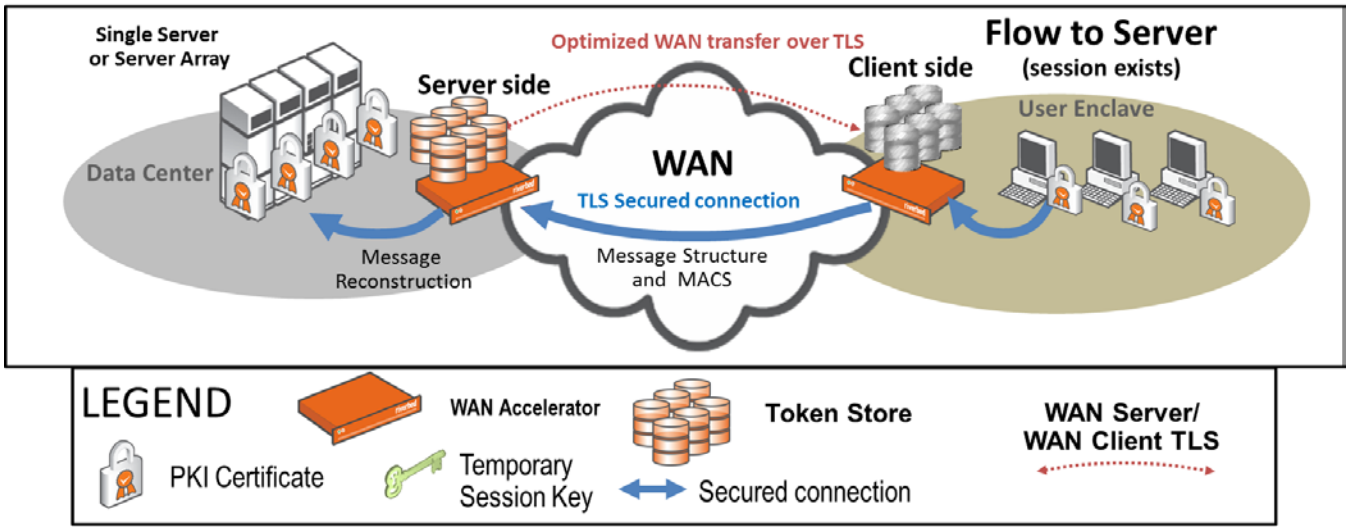

Fig. 7. Flow from Client Side

V. Flows In A High Integrity System

The overall recommended flows for the high assurance WAN Accelerator on the server-side are provided in Fig. 6. The TLS session keys are passed to the server-side WAN accelerator as shown in Fig. 3 and subsequently passed on to the client-side accelerator through the WAN Server / WAN Client TLS. The primary addition is the transmission of the original message fragmentation, padding, and MACs. These are needed to make an identical reconstruction of the original message. The original MACs can be added to the reconstructed message, for which the targeted end-point has the MAC key.

The original TLS MAC digests, paddings, and fragmentation are used to construct TLS packets in lieu of the values that would normally be computed by the TLS protocol. The client-side flows are given in Fig. 7 and show the same transmission of the fragmentation, paddings, and MACS for the reconstruction of the message on the server-side.

To ensure end-to-end integrity it is necessary that the MAC secret is not available to the WAN accelerator. The receiving WAN accelerator node uses the original content, as received and reconstructed from the WAN connection, the original TLS fragmentation information, and the associated MAC and padding values to construct valid TLS messages to the receiving endpoint. Because only the sending endpoint has the MAC secret, the receiving endpoint has an end-to-end integrity guarantee.

VI. Summary

We have reviewed the basic approaches to WAN Acceleration security when dealing with encrypted traffic. These have been found lacking in the specific areas of key protection, and message integrity. Key protection is lacking in that out-of-band passing of the private keys of the servers violate specific security tenants. Message integrity is lacking in that piecewise integrity is substituted for overall end-to-end integrity. We have also described the high assurance architectures and protection elements they provide. In order to preserve the high assurance key security and integrity elements, changes to the basic flows must be made. Key security can be maintained by passing only the session key which provides a lower risk than passing the private key of the server and maintains the unique identity of the server. Overall message integrity can be maintained by conspicuously reconstructing the original messages, including the message authentication codes. The development of WAN accelerator mechanisms does not require distribution of a private key which is often done with today's appliances. The distribution of private keys is a fundamental violation of a high assurance model. What remains is the need for high reliability and secure code for passing of session keys, as well as secure means of transporting TLS plans and MACs for the establishment of service interfaces on the appliances. Pilots are in the process of being undertaken, and we expect some modifications as we learn more about how these devices work.

This research is part of a body of work for high assurance enterprise computing using web services. Elements of this work include bi-lateral end-to-end authentication using PKI credentials for all person and non-person entities, a separate SAML credential for claims based authorization, fully encrypted at the transport layer and a defined federation process. Many of the elements of this work are described in [22-35].

References

[1] Rouse, Margaret, "WAN accelerator," Network management and monitoring: The evolution of network control, http://searchenterprisewan.techtarget.com/definition/WAN-accelerator, December 2009.

[2] Rouse, Margaret, "WAN optimization (WAN acceleration)," http://searchenterprisewan.techtarget.com/definition/WAN-optimization, August 2010.

[3] Machowinski, Matthias. "WAN optimization market passes $1 billion in 2008, up 29%; enterprise router market down." Enterprise Routers and WAN Optimization Appliances. Infonetics Research. Retrieved 19 July 2011.

[4] Skorupa, Joe; Severine Real (2010). "Forecast: Application Acceleration Equipment, Worldwide, 2006–2014, 2Q10 Update." Gartner, Inc. Retrieved 19 July 2011.

[5] Cardwell, N.; Savage, S.; Anderson, T. "Modeling TCP latency." INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE. Dept. of Comput. Sci. & Eng., Washington Univ., Seattle, WA: IEEE.org. Retrieved 20 July 2011.

[6] Jacobson, Van. "TCP Extensions for Long-Delay Paths." Request for Comments: 1072. Internet Engineering Task Force (IETF). Retrieved 19 July 2011.

[7] Floyd, Sally. "HighSpeed TCP for Large Congestion Windows." Request for Comments: 3649. Internet Engineering Task Force (IETF). Retrieved 19 July 2011.

[8] Paris, Chandler. "Latency & Colocation.". Retrieved 20 July 2011.

[9] Mark Rabinovich, Igor Gokhman. "CIFS Acceleration Techniques." Storage Developer Conference, SNIA, Santa Clara 2009.

[10] Conway, Richard (204). *Code Hacking: A Developer's Guide to Network Security*. Hingham, Massachusetts: Charles River Media. p. 281. ISBN 1-58450-314-9.

[11] Chang, Rocky (October 2002). "Defending Against Flooding-Based Distributed Denial-of-Service Attacks: A Tutorial." *IEEE Communications Magazine* **40** (10): 42–43

[12] Mehmet Altinel , Christof Bornhövd , Sailesh Krishnamurthy , C. Mohan , Hamid Pirahesh , Berthold Reinwald, Cache tables: paving the way for an adaptive database cache, Proceedings of the 29th international conference on Very large data bases, pp.718–729, September 09-12, 2003, Berlin, Germany

[13] Amiri, K., Tewari, R., Park, S., and Padmanabhan, S. 2002. On space management in a dynamic edge cache. In Proceedings of the Fifth International Workshop on the Web and Databases (WebDB 2002) (Madison, Wisc.). ACM, New York, 37–42.

[14] Jesse Anton , Lawrence Jacobs , Xiang Liu , Jordan Parker , Zheng Zeng , Tie Zhong, Web caching for database applications with Oracle Web Cache, Proceedings of the 2002 ACM SIGMOD international conference on Management of data, June 03-06, 2002, Madison, Wisconsin [doi>10.1145/564691.564762]

[15] CacheFlow. 1999. Accelerating e-commerce with CacheFlow internet caching appliances (a CacheFlow white paper).

[16] Cain, B., Spatscheck, O., May, M., and Barbir, A. 2001. Request-routing requirements for content internetworking. http://www.ietf.org/internet-drafts/draft-cain-request-routing-req-03.txt.

[17] K. Selçuk Candan , Wen-Syan Li , Qiong Luo , Wang-Pin Hsiung , Divyakant Agrawal, Enabling dynamic content caching for database-driven web sites, Proceedings of the 2001 ACM SIGMOD international conference on Management of data, p.532-543, May 21-24, 2001, Santa Barbara, California, USA [doi>10.1145/375663.375736]

[18] Erich Gamma , Richard Helm , Ralph Johnson , John Vlissides, Design patterns: elements of reusable object-oriented software, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, 1995

[19] Request for Comments: The Transport Layer Security (TLS) Protocol Version 1.2, http://tools.ietf.org/html/rfc5246, August 2008

[20] Request for Comments: 3749, Transport Layer Security Protocol Compression Methods, http://tools.ietf.org/html/rfc3749, May 2004

[21] Request for Comments: Transport Layer Security (TLS) Extensions, http://tools.ietf.org/html/rfc4366, April 2006.

[22] William R. Simpson, Coimbatore Chandersekaran and Andrew Trice, Electronic Digest of the 2008 System and Software Technology Conference, "A Persona-Based Framework for Flexible Delegation and Least Privilege," Las Vegas, Nevada, May 2008.

[23] William R. Simpson, Coimbatore Chandersekaran and Andrew Trice, The 1st International Multi-Conference on Engineering and Technological Innovation: IMET2008, "Cross-Domain Solutions in an Era of Information Sharing," Volume I, pp.313–318, Orlando, FL, June 2008.

[24] Coimbatore Chandersekaran and William R. Simpson, World Wide Web Consortium (W3C) Workshop on Security Models for Device APIs, "The Case for Bi-lateral End-to-End Strong Authentication," 4 pp., London, England, December 2008.

[25] William R. Simpson and Coimbatore Chandersekaran, The 2nd International Multi-Conf.on Engineering and Technological Innovation: IMETI2009, Volume I, pp. 300–305, "Information Sharing and Federation", Orlando, FL, July 2009.

[26] Coimbatore Chandersekaran and William R. Simpson, The 3rd International Multi-Conf. on Engineering and Technological Innovation: IMETI2010, Volume 2, "A SAML Framework for Delegation, Attribution and Least Privilege," pages 303–308, Orlando, FL., July 2010.

[27] William R. Simpson and Coimbatore Chandersekaran, The 3rd International Multi-Conference on Engineering and Technological Innovation: IMETI2010, Volume 2, "Use Case Based Access Control," pages 297–302, Orlando, FL, July 2010.

[28] Coimbatore Chandersekaran and William R. Simpson, The First International Conference on Computer Science and Information Technology (CCSIT-2011), "A Model for Delegation Based on Authentication and Authorization," Springer Verlag Berlin-Heildleberg, Lecture Notes in Computer Science 20 pp.

[29] William R. Simpson and Coimbatore Chandersekaran, The 16th International Command and Control Research and Technology Symposium: CCT2011**,** Volume II, pp. 84–89**,** "An Agent Based Monitoring System for Web Services," Orlando, FL, April 2011.

[30] William R. Simpson and Coimbatore Chandersekaran, International Journal of Computer Technology and Application (IJCTA)**,** "An Agent-Based Web-Services Monitoring System" Vol. 2, No. 9, September 2011, pp. 675–685.

[31] William R. Simpson, Coimbatore Chandersekaran and Ryan Wagner, Lecture Notes in Engineering and Computer Science, Proceedings World Congress on Engineering and Computer Science 2011, Volume I, "High Assurance Challenges for Cloud Computing," pp. 61–66, San Francisco, October 2011.

[32] Coimbatore Chandersekaran and William R. Simpson, Lecture Notes in Engineering and Computer Science, Proceedings World Congress on Engineering 2012, The 2012 International Conference of Information Security and Internet Engineering, Volume I, "Claims-Based Enterprise-Wide Access Control," pp. 524–529, London, July 2012.

[33] William R. Simpson and Coimbatore Chandersekaran, Lecture Notes in Engineering and Computer Science, Proceedings World Congress on Engineering 2012, The 2012 International Conference of Information Security and Internet Engineering, Volume I, "Assured Content Delivery in the Enterprise," pp. 555–560, London, July 2012.

[34] William R. Simpson and Coimbatore Chandersekaran, Lecture Notes in Engineering and Computer Science, Proceedings World Congress on Engineering and Computer Science 2012, Volume 1, "Enterprise High Assurance Scale-up," pp. 54–59, San Francisco, October 2012.

[35] Coimbatore Chandersekaran and William R. Simpson, International Journal of Scientific Computing, Vol. 6, No. 2, "A Uniform Claims-Based Access Control for the Enterprise," December 2012, ISSN: 0973-578X, pp. 1–23.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE (DD-MM-YY)	2. REPORT TYPE	3. DATES COVERED (From – To)
03-07-15	Non-Standard	

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Wide Area Network Acceleration in a High Assurance Enterprise	HQ0034-14-D-0001
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBERS

6. AUTHOR(S)	5d. PROJECT NUMBER
William R. Simpson, Kevin E. Foltz	BC-5-2283
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESSES	8. PERFORMING ORGANIZATION REPORT NUMBER
Institute for Defense Analyses 4850 Mark Center Drive Alexandria, VA 22311-1882	NS D-5404 H15-000020

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR'S / MONITOR'S ACRONYM
Frank P. Konieczny USAF HQ USAF SAF/CIO A6	SAF/CIO/CTO
	11. SPONSOR'S / MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Project Leader: William R. Simpson

14. ABSTRACT

Bandwidth continues to be a problem for the active enterprise. One solution to bandwidth problems over long-haul distances with restricted bandwidth is the Wide Area Network (WAN) Accelerator. This accelerator works by tokenizing blocks of information that are sent multiple times in network traffic. Because many such communications include previously transmitted material, the accelerator traffic quickly damps out to transmissions that include tokens instead of the original communication. The tokens are reconstituted before delivery, and the receiver has a seamless connection and is unaware of the process. The acceleration is not without its drawbacks. The process does not work on encrypted traffic due to the random nature of encryption. For high assurance systems using an end-to-end paradigm, there are two main areas of concern. The first is security (how do we handle the decryption/re-encryption process?). The second is integrity (how do we maintain end-to-end integrity when encryption is broken?).

This paper discusses the current approach to WAN acceleration and the changes that are required by a high assurance end-to-end approach. The latter rely on a well-formed security paradigm for the enterprise.

15. SUBJECT TERMS

Network Acceleration Appliance, Protection, IT Security, Encryption, Key Management, Wide Area Network, Integrity

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Frank P. Konieczny
Unclassified	Unclassified	Unclassified	Unlimited	6	19b. TELEPHONE NUMBER (Include Area Code) (703) 697-1308