ETSI TR 103 421 V1.1.1 (2017-04)

CYBER;
Network Gateway Cyber Defence

Reference
DTR/CYBER-0015

Keywords
cyber security, information assurance, privacy

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

The present document can be downloaded from:
http://www.etsi.org/standards-search

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the only prevailing document is the print of the Portable Document Format (PDF) version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status.
Information on the current status of this and other ETSI documents is available at
https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx

If you find errors in the present document, please send your comment to one of the following services:
https://portal.etsi.org/People/CommiteeSupportStaff.aspx

Copyright Notification

No part may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm except as authorized by written permission of ETSI.
The content of the PDF version shall not be modified without the written authorization of ETSI.
The copyright and the foregoing restriction extend to reproduction in all media.

© European Telecommunications Standards Institute 2017.
All rights reserved.

DECT™, **PLUGTESTS**™, **UMTS**™ and the ETSI logo are Trade Marks of ETSI registered for the benefit of its Members.
3GPP™ and **LTE**™ are Trade Marks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners.
GSM® and the GSM logo are Trade Marks registered and owned by the GSM Association.

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

Foreword

This Technical Report (TR) has been produced by ETSI Technical Committee Cyber Security (CYBER).

Modal verbs terminology

In the present document "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Executive summary

The present document provides an overview and recommendations concerning cyber defence capabilities at network gateways. The capabilities are implemented using what are usually referred to as "middleboxes" that may be integrated into traffic routers that typically exist at boundaries between networks. Network gateways are critically important points for implementing cyber defence in conjunction with other essential functions.

The present document notes that network gateway cyber defence related standards activities have increased significantly because of an array of use cases combined with the rapidly increasing encryption of traffic occurring between end points where network application servers are interacting directly with software clients on end user devices. The use cases consist of an array of business and compliance obligations. The present document then continues to derive a set of related cyber defence technical requirements that include:

1) secure and controlled exposure of traffic observables;

2) sufficient observable information for acquisition and analysis for defence measures; and

3) the ability to institute defence measures as part of gateway management.

The present document then examines the emerging new challenges and mechanisms for gateway cyber defence. The challenges include virtualization implementations, 5G mobile systems, Internet of Things deployments, Over The Top services, and "encrypt everything" initiatives. On the positive side, the considerable industry and academic research and development efforts have produced a combination of existing protocol adaptations and effective new protocols and platforms that have considerable promise - especially one known as mcTLS.

The present document concludes with several recommendations that include a consensus view on what information and secure access capabilities are required to support gateway cyber defence, what steps the ETSI Cybersecurity Technical Committee should take for a new Technical Specification to support the requirements, and how collaboration with external bodies might encourage use of gateway cyber defence capabilities.

Introduction

A network gateway is a device that enables or facilitates the interconnecting of networks or applications via those networks. They have existed since the origins of electronic communication. With the emergence of packet data networks, they have assumed many different roles, including cyber defence. Those additional roles are commonly denominated as "middlebox" functions [i.3]. An especially common network gateway used for cyber defence purposes is referred to as a *firewall* - defined by 3GPP as a functional entity which blocks or permits the flow of various traffic types based on a set of policy rules and definitions. All signalling to internal network resources can be directed via a network gateway dedicated to that purpose.

Network gateways serve many critical needs that include management of network traffic and meeting service level agreement or regulatory requirements. One of those critical needs is that of cyber defence - which can be met through the detection and prevention of threats at the external border point of all kinds of networks ranging from a national infrastructure to an organization or home network. Deep Packet Inspection capabilities are widely deployed to facilitate these capabilities. However, the appearance of ever more sophisticated threats and adaptive malware is proving challenging to detection and blocking efforts.

A significant cyber security challenge emerging today is the combination of Over the Top services combined with "encrypt everything" initiatives that generated potentially huge amounts of traffic between some arbitrary service portal somewhere in the world, and an end user's terminal - even an application on a device. Some Internet of Things implementations also fall into this category. While these steps meet significant needs today, these practices may have adverse effects such as impeding detection of malware and other cyber security threats, as well as managing network traffic and meeting a broad array of business, organizational, and regulatory requirements. A balanced approach is needed that provides support to all the requirements that exist today.

The emergence of NFV-SDN implementations is engendering considerable new efforts to virtualize network gateway capabilities. These efforts include the use of on-demand Big Data Analysis to more rapidly detect and mitigate threats.

Many different industry forums today are examining network gateway requirements and solutions available - largely as insular work items and projects. The present document assembles an understanding of the related ecosystem, models, protocols, and implementation mechanisms for gateway-based cyber defence.

1 Scope

The present document provides an overview and recommendations concerning cyber defence capabilities at network gateways. It analyses the network gateway cyber defence ecosystem, technical requirements, new challenges and techniques and then draws recommendations for new standardization work in that area.

2 References

2.1 Normative references

Normative references are not applicable in the present document.

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication ETSI cannot guarantee their long term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] SIGCOMM '15, Naylor et al., Multi-Context TLS (mcTLS): "Enabling Secure In-Network Functionality in TLS", August 17 - 21, 2015, London, United Kingdom.

NOTE: Available at http://conferences.sigcomm.org/sigcomm/2015/pdf/papers/p199.pdf.

[i.2] ETSI TR 103 456: "CYBER; Implementation of the Network and Information Security (NIS) Directive".

[i.3] IETF RFC 3224: "Middleboxes: Taxonomy and Issues", February 2002.

[i.4] IETF draft-mm-wg-effect-encrypt-04: "Effect of Ubiquitous Encryption", October 2016.

[i.5] OASIS CybOX™ Version 2.1.1. Part 01: "Overview".

NOTE: Available at http://docs.oasis-open.org/cti/cybox/v2.1.1/cybox-v2.1.1-part01-overview.pdf. See also, CybOX Project/specifications, https://github.com/CybOXProject/specifications/wiki.

[i.6] NIST SP 800-117: "Guide to Adopting and Using the Security Content Automation Protocol (SCAP)".

[i.7] SP 800-126 Revision 2: "The Technical Specification for the Security Content Automation Protocol (SCAP)".

[i.8] Recommendation ITU.T X.1500: "Overview of cybersecurity information exchange".

NOTE: See http://www.itu.int/itu-t/recommendations/rec.aspx?rec=11060.

[i.9] IETF RFC 7632: "Endpoint Security Posture Assessment: Enterprise Use Cases".

[i.10] IETF draft-ietf-sacm-requirements-15: "Security Automation and Continuous Monitoring (SACM) Requirements".

NOTE: Available at https://datatracker.ietf.org/doc/draft-ietf-sacm-requirements/.

[i.11] ETSI TS 101 331 (V1.4.1): "Lawful Interception (LI); Requirements of Law Enforcement Agencies".

[i.12] ETSI TS 133 106 (V13.4.0): "Universal Mobile Telecommunications System (UMTS); LTE; 3G security; Lawful interception requirements (3GPP TS 33.106 version 13.4.0 Release 13)".

[i.13] ETSI TS 102 656 (V1.2.2): "Lawful Interception (LI); Retained Data; Requirements of Law Enforcement Agencies for handling Retained Data".

[i.14] Recommendation ITU-T X.1038: "Security Requirements and reference architecture for Software-Defined Networking" (10/2016).

NOTE: Available at http://www.itu.int/itu-t/recommendations/rec.aspx?rec=13058.

[i.15] 5G PPP Architecture Working Group, View on 5G Architecture, Version 1.0, July 2016.

NOTE: Available at https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-5G-Architecture-WP-July-2016.pdf.

[i.16] European Commission, Copyright and Neighbouring Rights.

NOTE: Available at , http://ec.europa.eu/internal_market/copyright/documents/index_en.htm.

[i.17] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications).

[i.18] StepExchange: "Combatting nuisance calls and texts".

NOTE: Available at https://www.stepchange.org/Portals/0/documents/media/reports/additionalreports/Designed_nuisance_calls_appendix_final.pdf.

[i.19] ENISA: "Resilience and security of communication infrastructure, networks and services".

NOTE: Available at https://resilience.enisa.europa.eu/.

[i.20] CRS: "National Security and Emergency Preparedness Communications", 19 September 2012.

NOTE: Available at https://fas.org/sgp/crs/natsec/R42740.pdf.

[i.21] ACM: "Proceedings of the 2016 ACM Workshop on Cloud-Assisted Networking".

NOTE: Available at http://dl.acm.org/citation.cfm?id=3010079.

[i.22] Securebox: Toward Safer and Smarter IoT Networks.

NOTE: Available at https://www.cs.helsinki.fi/u/yding/publications/securebox-pre-camera.pdf.

[i.23] Embark: "Securely Outsourcing Middleboxes to the Cloud".

NOTE: Available at http://forum.stanford.edu/events/2016/slides/iot/Chang.pdf.

[i.24] Draft Recommendation ITU-T Y.gw-IoT-arch: "Functional architecture of gateway for IoT applications".

[i.25] Draft Recommendation ITU-T Y.IoT-cdn: "Framework of constrained-device networking in the IoT environments".

[i.26] ResearchGate: "Impact of Over the Top (OTT) Services on Telecom Service Providers".

NOTE: Available at https://www.researchgate.net/publication/276175550_Impact_of_Over_the_Top_OTT_Services_on_Telecom_Service_Providers.

[i.27] MarketsandMarkets: "Over the Top Market by Content Type, by Platform (Smart Devices, Laptops, Desktops, and Tablets), by Service (Consulting, Installation, and Maintenance), by Revenue Model, by Deployment Model, by Vertical, by User Type, by Region - Global Forecast to 2020".

NOTE: Available at http://www.marketsandmarkets.com/Market-Reports/over-the-top-ott-market-41276741.html.

[i.28] IAB: "The effect of encrypted traffic on the QoS mechanisms in cellular networks".

NOTE: Available at https://www.iab.org/wp-content/IAB-uploads/2015/08/MaRNEW_1_paper_25.pdf.

[i.29] BuiltWith®: "SSL by Default Usage Statistics".

NOTE: Available at https://trends.builtwith.com/ssl/SSL-by-Default.

[i.30] IAB: "Managing Radio Networks in an Encrypted World (MaRNEW) Workshop 2015".

NOTE: Available at https://www.iab.org/activities/workshops/marnew/.

[i.31] Fraunhofer FKIE: "White Paper on Encrypted Traffic Management", January 2016.

NOTE: Available at http://images.machspeed.bluecoat.com/Web/BlueCoat/%7Bab90f902-7c34-440d-a933-0a33f59718ce%7D_20160421-ETM-Paper_English_final.pdf.

[i.32] Layer 9: "Session 4 (Middleboxes) -- Paper 1: Multi-Context TLS (mcTLS): Enabling Secure In-Network Functionality in TLS".

NOTE: Available at http://www.layer9.org/2015/08/session-4-middleboxes-paper-1-multi.html.

[i.33] Intel®: "Intel® Software Guard Extensions (Intel® SGX)".

NOTE: Available at https://software.intel.com/en-us/sgx.

[i.34] Recommendation ITU-T X.1542: "Session information message exchange format".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply:

firewall: functional entity which blocks or permits the flow of various traffic types based on a set of policy rules and definitions

middlebox: any intermediary box performing functions apart from normal, standard functions of an IP router on the data path between a source host and destination host, including network gateways [i.3]

network gateway: device or system that enables or facilitates the interconnecting of networks or applications via those networks

observable: described definitive characteristic of an object observed in the cyber environment that facilitates a common structure relating to the specification, capture, characterization and communication of events [i.5]

3.2 Abbreviations

For the purposes of the present document, the following abbreviations apply:

3GPP	Third Generation Partnership Project
5GPPP	5G Infrastructure Public Private Partnership
ACM	Association for Computing Machinery
ALPN	Application-Layer Protocol Negotiation

API	Application Program Interface
ARF	Asset Reporting Format
AT-TLS	Application Transparent Transport Layer Security
ATTM	Access, Terminals, Transmission and Multiplexing committee
CA	Certification Authority
CAGR	Compound Annual Growth Rate
CBOR	Concise Binary Object Representation
CCSS	Common Configuration Scoring System
CN	Core Network
CPE	Common Platform Enumeration
CRS	Congressional Research Service
CTI	Cyber Threat Intelligence
CYBEX	Cybersecurity Information Exchange
CybOX	Cyber Observable Expression
DLP	Data Loss Prevention
DMCA	Digital Millennium Copyright Act
DPI	Deep Packet Inspection
ECMA	European Computer Manufacturers Association
ECN	Explicit Congestion Notification
ENISA	European Union Agency for Network and Information Security
ETI	Encrypted Traffic Inspection
EV	Extended Validation
FKIE	Fraunhofer Institute for Communication, Information Processing and Ergonomics
GSM	Global System for Mobile communication
GSMA	GSM Association
HICCUPS	Handshake-based Integrity Check of Critical Underlying Protocol Semantics
HTTP	HyperText Transfer Protocol
HTTPS	HTTP Secure (also HTTP over TLS)
IAB	Internet Architecture Board
IACD	Integrated Adaptive Cyber Defence fabrics
ID	Identity
IDS	Intrusion Detection System
IETF	Internet Engineering Task Force
IoT	Internet of Things
IP	Internet Protocol
IPR	Intellectual Property Requirements
IPS	Intrusion Prevention Systems
ISG	Industry Specification Group
ISP	Internet Service Provider
ITU-T	International Telecommunication Union-Telecommunication Standardization Sector
LTE	Long Term Evolution
MAEC	Malware Attribute Enumeration and Characterization
MAMI	Measurement and Architecture for a Middleboxed Internet
MARCOM	Marketing and Communications
mbTLS	Middlebox Transport Layer Security
mcTLS	Multi-Context Transport Layer Security
MITM	Man In The Middle
MNO	Mobile Network Operator
MSP	Multihoming Service Provider
NAT	Network Address Translation
NFV	Network Functions Virtualisation
NFV-SDN	Network Functions Virtualisation-Software Defined Networks
NGFW	Next Generation FireWalls
NIS	Network and information systems
OASIS	Organization for the Advancement of Structured Information Standards
OpenC2	Open Command and Control
OS	Operating System
OSP	Online Service Provider
OTT	Over The Top
OVAL	Open Vulnerability and Assessment Language
PPP	Public Private Partnership
RAN	Radio Access Network

RAR	Rotate And Release
RFC	Request For Comments
SACK	Selective Acknowledgment
SACM	Security Automation and Continuous Monitoring
SCAP	Security Content Automation Protocol
SCP	Smart Card Platform
SDN	Software Defined Network
SEMI	Stack Evolution in a Middlebox Internet
SGX	Software Guard Extensions
SIMEF	Session Information Message Exchange Format
SPAN	Services and Protocols for Advanced Networks
SPUD	Substrate Protocol for User Datagrams
SSL	Secure Sockets Layer
STIX	Structured Threat Information eXchange
TCP	Transmission Control Protocol
TG	Throughput Guidance
TGK	Telekommunikationsgesetz (Telecommunications Law)
TIPHON	Telecommunications Internet Protocol Harmonization Over Networks
TISPAN	Telecommunications and Internet converged Services and Protocols for Advanced Networking
TLS	Transport Layer Security
TLS-AUX	Transport Layer Security Auxiliary Data
TLS-RaR	Transport Layer Security Rotate and Release
TMSAD	Trust Model for Security Automation Data
UDP	User Datagram Protocol
UE	User Entity
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
USD	US Dollars
VM	Virtual Machine
WAN	Wide Area Network
XCCDF	Extensible Configuration Checklist Description Format

4 Network gateway cyber defence ecosystem: activities and use cases

4.1 Introduction - the gateway as a protection element

This clause provides an overview of the gateway as a protection element by describing the diverse standards activities occurring in industry bodies as well as gateway cyber defence business and compliance obligation use cases.

As stated in the introduction a system may be protected by a firewall that exposes the system through managed entry points. The normal visualization of a wall is misleading and in fact a more realistic visualization is that of an enclosing sphere, with the entry point, the gateway, being the only access to the protected domain. Thus, for a gateway to work the core assertion for security is that the gateway is the only access point to the protected domain.

A gateway as the point of access to the internal network may need to prevent access to hostile users, traffic and content. In order to achieve this, the capabilities at the gateway are necessarily broad in scope and deep in terms of protocol stacks. For example, protection against malicious payloads may require that file transfers are cached at the gateway and examined to identify the presence of viruses or Trojan horses and similar. If a protected domain is subject to a Denial of Service attack it may be necessary to distribute the gateway itself away from the protected domain.

A gateway may act to police traffic leaving the protected domain in addition to policing traffic entering the protected domain. This may be achieved using the same techniques in each direction. Port filtering of IP packets has been used but such practices can be bypassed, and if this is done, the bypass should not result in a security leak. Thus, middlebox techniques to ensure safe and secure firewall traversal may be required to be implemented at the gateway.

4.2 Network gateway cyber defence related standards activities

A considerable array of diverse industry and intergovernmental standards bodies are today considering the requirements and solutions for network gateway cyber defence. This ecosystem is depicted in Figure 1.

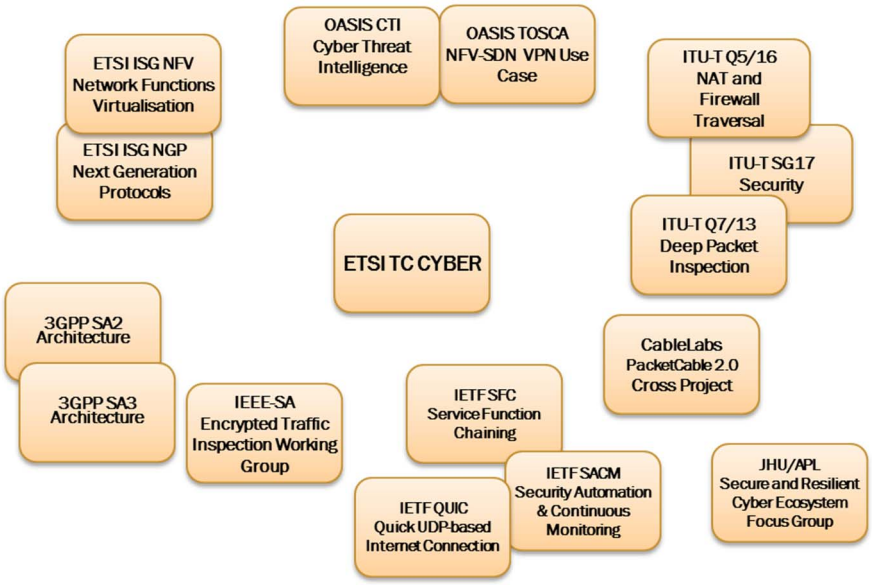


Figure 1: Organizations and groups treating gateway cyber defence

In the past several years, the amount of research and development is reflected in a rather considerable number of papers, presentations, and initial standards work occurring in almost every network security venue - especially under the generic term "middlebox security". Twenty of the more prominent advocated techniques are summarized in Clause 6.3.

4.3 Network gateway cyber defence business and compliance obligation use cases

4.3.1 Cyber security use cases

Gateways are routinely used by enterprise and some citizens to control what data goes in and out of a network and provide for data security and cyber security functionality. These gateways have continuing requirements, and need to complement, not break, security. The security gateway uses include scanning traffic, searching for threats (e.g. malware) or data leakage, doing single sign on, changing traffic to enforce company's policy (e.g. replace virus by block message), protecting all kinds of clients and prevent misbehaving devices from attacking networks (including IoT).

Some enterprise networks have highly stringent cyber security requirements because they are carrying information regarded as having considerable value. Such enterprises would typically be found across the sectors of government, finance, petrochemical, pharmaceutical and technology, although any business could consider its information worthy of high protection. These enterprises need a path that puts them in control of their cyber security defences, so they can determine what visibility and access each defensive component has. This may be on a granular level depending on context: say, some whitelisted information flows are considered trusted end-to-end by defensive gateways and thus only minimal supporting cyber defensive data need be collected on them, whereas others may require full inspection and authorization in- or out- of a gateway. The important point is that the owner of the enterprise should be in control, so can make the appropriate decisions for their usage. Middleboxes have become frequently used to meet these requirements in the face of increasing deployment of Transport Layer Security (TLS) protocols as described in Clause 6.

Forensics is a critical component of cyber security. Once a breach is identified, investigators determine how the information at the gateway helps. Often it is hard to associate sessions seen by the gateway with an application program/software/device at the communication end point. Logging references that associate the device view with the gateway view would be an immediate benefit and would strike a privacy balance.

4.3.2 Network management use cases

Denial of service or denial of network performance are significant drivers for protection. Gateways that can detect and block the offending sessions can be very effective. However they need sufficient visibility of the addresses and types of content in order to do this without blocking everything. These needs also apply to content delivery networks which also have a cognisance role and awareness of their state and cyber security controls.

These requirements necessitate servers share information with gateways. It is not just about client endpoints and their enterprise gateways. The information might include filenames and file hashes - so that gateways can be assured (to some extent) that the content being passed under encrypted tunnels does not contain malware or blocked files such as illegal explicit images. These could be used for incident forensics after a new strain of malware and deployment mechanism is identified.

There is considerable value to a network services provider (ISP or MNO) in providing cyber security services to their customers. This includes what can they do with a network or cloud hosted gateway, what information does this need to be really effective. For example, a citizen might appoint a security company to run a cloud gateway cyber security protection service - they would then route all their IoT WiFi and Cellular traffic through this cloud gateway and authorize it to have extra session description metadata and block/filter/manage routing.

As the network infrastructures (ISP or MNO) carry increasing volumes of critical traffic perhaps increasingly opaquely, the network infrastructure itself becomes a super-critical infrastructure, and thus part of Critical National Infrastructures. Therefore national requirements for service continuity, resilience, identification and recovery can be expected. The cyber defensive gateways of the networks are thus the natural places to deliver many of these requirements, and thus in order to safely and securely fulfil these functions they will require a minimal level of information about network flows they are carrying.

Other network management functions of middleboxes include:

- Routing related: route/traffic shaping/QoS network traffic based on properties of network flow content (e.g. URL of HTTP request).
- Caching: Device caches content to reduce used bandwidth.
- Data Compression Proxy: Device reduces used network bandwidth by compressing content (e.g. gzip or changing image quality).
- WAN Optimization.

4.3.3 Device and application management - discovery and health attestation use cases

Gateways discover what endpoints are on the connected networks and convey sufficient information to help with categorizations to support cyber security decisions, and be assured that the endpoints are in good health and in a known good state. Health status includes determination of a safe assured boot sequence, or information about OS/software versions and update/patch status. Likewise, if an endpoint determines which gateways have authority over them, then it can be trusted to share session description metadata. Sensitive communications require a mechanism to decide whether the gateway middlebox can be trusted.

These capabilities were initially implemented using the Security Content Automation Protocol (SCAP) and subsequently become part of security requirements such as NIST SP 800-117 [i.6], and Special Publication SP 800-126 Revision 2 [i.7]. SP 800-126 [i.7] in turn invokes an array of additional specifications including *Asset Reporting Format (ARF)*, Asset Identification, Common Configuration Scoring System (CCSS), and Trust Model for Security Automation Data (TMSAD), which provides support for digitally signing SCAP source and result content. SCAP also includes support for the *Open Vulnerability and Assessment Language (OVAL), Common Platform Enumeration (CPE)*, and *Extensible Configuration Checklist Description Format (XCCDF)*. All of these capabilities have significant dependencies with gateway defence middleboxes. Most of these specifications are also part of the ITU-T Cybersecurity Information Exchange (CYBEX) specification, Recommendation ITU-T X.1500 [i.8] and been republished by ITU-T as part of the CYBEX suite.

Some of the requirements and technical capabilities have also been undertaken by the IETF Security Automation and Continuous Monitoring (SACM) Working Group. See IETF RFC 7632 [i.9]. See also, *Security Automation and Continuous Monitoring (SACM) Requirements*, draft-ietf-sacm-requirements-15 [i.10].

4.3.4 Industry specifications and agreement use cases

Communication systems require the cooperation among a large number of network operators who cooperate in transporting the traffic between end points without endangering the networks and end users. This cooperation is maintained through the development and application of specifications and entering into service agreements that require the traffic meet certain requirements, such it be free from malware or spoofed identities.

The GSMA Fraud and Security Group, for example, has published numerous specifications, guides, and agreements highly dependent on middlebox gateway defence capabilities used by mobile operators worldwide to enhance security and reduce risk in mobile networks and also leverage related specifications in 3GPP SA technical committees.

4.3.5 Lawful interception and retained data use cases

Lawful Interception and retained data are invoked in both public and private networks by combinations of legal requirements aimed at providing, when directed by lawful government authority, forensics at network gateway middleboxes that are essential for investigations and pursuit of criminal and civil remedies. In general, the applicable law and technical requirements in jurisdictions require unencrypted access to these network forensics. See, e.g. ETSI TS 101 331 [i.11], ETSI TS 133 106 [i.12] and ETSI TS 102 656 [i.13].

4.3.6 Intellectual property protection use cases

In many instances, gateway devices enforce document permissions and privileges and traffic availability for intellectual property protection purposes. Many products exist which facilitate enforcement through analysis of traffic at gateway middleboxes. There are many such examples relied upon by intellectual property owners and their industry organization representatives. These mechanisms require access to either the full content or a trusted inspection mechanism that can meet the use cases at defence gateways. In most cases, because the protected intellectual property consists of large files or high bandwidth streaming and the distribution multi-threaded, the connectivity service providers have a significant interest in managing their networks to prevent them being adversely affected by such activities.

In the U.S. the Digital Millennium Copyright Act (DMCA) provides a mechanism for copyright holders to protect their online content. Once an author becomes aware that his work is being infringed upon it is up to the copyright holder to notify the Internet Service Provider (ISP)/Online Service Provider (OSP) that contains the material about the alleged infringement. If the ISP/OSP has created a mechanism that enables copyright holders to request that infringing work be removed and the process is followed by the ISP/OSP, the DMCA generally provides the ISP/OSP with a safe harbour from liability.

Intellectual property law in Europe varies by country. In addition to the Berne Convention implementations within individual countries, European Copyright law is promulgated via directives. Intellectual property directives provide member states guidance on how to regulate Internet and electronic media copyright issues, and are available at: European Commission, Copyright and Neighbouring Rights [i.16].

There is therefore a requirement for some networks to understand the copyrighted material transiting their gateways.

4.3.7 End user privacy and protection of minors

The ability to protect end user privacy by detecting and preventing unwanted eMail, text, fax, and voice communication depends on the ability of gateway middlebox detection and halting of such traffic so it does not proceed to the end user. Increasingly, national, regional, and international law require collaboration of providers at gateways to identify the traffic and prevent its further transmission. See e.g. the Directive 2002/58/EC [i.17]); TGK as implemented by 58BNetzA 2013d: Themenblatt Rufnummernmissbrauch; U.S. Telephone Consumer Protection Act. See generally, StepExchange, Combatting nuisance calls and texts [i.18].

Related use cases include parental or public institutions (such as schools, public libraries, religious organizations) filtering of inappropriate or unlawful content commonly referred to as parental filtering. In all of these use cases, communication with gateway middleboxes is necessary to detect and manage the traffic. An important class of these use cases is the domestic gateway for parental filtering.

4.3.8 Resilience and security of communication infrastructure, networks and services

The implementation of an array of essential communications and resiliency during times of declared widespread emergency as well as the protection of critical infrastructure is critically dependent on the ability to characterize, prioritize, and otherwise manage traffic at network gateway middleboxes. A critical use case is to be able to rapidly identify and isolate an entire class of traffic that is malfunctioning or being used for a widespread cyber attack. This requires sufficient visibility down the stack of the device types under transmission. See ENISA, Resilience and security of communication infrastructure, networks and services [i.19]. See also CRS, National Security and Emergency Preparedness Communications [i.20].

Denial of service attacks - including those on systems necessary to manage critical infrastructure - can emanate from widely distributed IoT devices located behind network gateways and underscore the bilateral symmetry of cyberdefence at these points - which requires detection and control of local network traffic to prevent attacks elsewhere.

5 Network gateway cyber defence technical requirements

5.1 Introduction

This clause describes technical requirements for gathering and sharing cyber defence information at gateways using middlebox capabilities. There is a growing body of related standardization work surveyed in clause 6. Typically, this work includes formats and specifications for information pertaining to a device/sensor/endpoint to server/client/ endpoint session, rather than the decision making and analytic tools that are applied to them. The term "observables" is used here because of its increasing use in the cyber threat environment to describe an observed definitive characteristic of an object that facilitates a common structure relating to the specification, capture, characterization, and communication of events.

5.2 Secure and controlled exposure of traffic observables

The guiding principle is that the endpoint should be in control of what observables are made available to middleboxes and the level of access the middleboxes have to those observables. However, one also needs to consider the situation where the endpoint is compromised in which case no undue harm should come to the middlebox, and it should still be possible for the middlebox to protect, in some regard, the user of the compromised endpoint. Having observables generated at a different level of privilege to application data would go some way to providing this.

The endpoint is either client or server, and the authorization will depend on the use-case but can be with either end or both together. For example, for enterprise cyber security, the authorization would be with the client; and by contrast for caching and content delivery, may only be with the server but could be with both client and server.

The authorizing endpoint or -points need a means to authenticate the middlebox and determine whether they trust it for the information and capability they are going to make available to it. This authentication would allow the endpoint to be sure of the ownership of the middlebox, and to what level of privileged access the middlebox can have, and may also allow the endpoint to attest that the middlebox is in secure condition and is operating as expected. This allows the endpoint to determine what observables it might expose to the middlebox, based on its policies with respect to the privileges presented. The attestation, where provided, is a means for the endpoint to determine how secure the middlebox is, again used as part of the policy for what to expose.

A general expectation would be that the middlebox was protecting against cyber security attacks at least as well as an endpoint. By virtue of an enterprise security middlebox having view of the cyber activity across an enterprise-worth of endpoints, it would have a more sophisticated view of cyber security within that network than any other single endpoint, and therefore should be capable of having a higher level of cyber security, and by its authority, extending that security protection to the endpoint. In any case, the endpoint has a determination of how secure the middlebox is, and if this is regarded as lower than it prefers, it would have policies which restrict the observable which it would expose to that middlebox. The middlebox may, in response, limit the level of external access available to the endpoint.

Table 1: Requirements for secure and controlled exposure of traffic observables

Requirement	Notes
Authorization of middlebox	Client or server or both are required to authorize middlebox access to observables, depending on use-case or whether input to this decision matters to that endpoint.
Security of middlebox	Middlebox is expected to be protected against cyber security attacks at least as well as an endpoint. Where this is not the case, it is possible for the endpoint to be aware of this, and make a decision to expose fewer or no observables.
Awareness of endpoints	It is expected that in most use-cases, both endpoints should be able to determine the presence of a middlebox although they may choose to ignore it.
Visibility of security contexts	It is expected that in most use-cases, an endpoint can determine the level of security protection on associated middlebox connections that it does not participate in directly. E.g. forwarding secure tunnel.
Compatibility with legacy TLS proxies	It may be required to connect with endpoints that are not conformant with middlebox security protocols. These should be able fall back to legacy TLS proxies but should still provide middlebox security assurances to the endpoint that is conformant.

5.3 Sufficient observable information for acquisition and analysis for defence measures

The middlebox should establish a secure connection with endpoints and receive observables. These should be trusted for their use to be effective, the middlebox can place a level of trust on them in order to handle the situation where the endpoint is itself compromised; this should be expected, so the actions taken should be commensurate with this expectation. If there is any endpoint performance expectation on the middlebox, such as the middlebox forwarding the connection with an agreed latency, then the middlebox should have a means of recording these performance expectations and should be able to indicate whether it can achieve them, i.e. can its buffering and analysis be done within the limit. If not, it would be sensible to agree a different performance expectation, which may require exposure of different observables. Means should be provided for this negotiation.

Identification of endpoints and connections/sessions is particularly critical so deserves special mention, as accurate recording improves cyber security analysis and retrospective forensics. Therefore, the middlebox would expect to be able to determine an identifier for each endpoint, that is meaningful of context and uniqueness. Similarly, for connections, it should be expected that the view of different middleboxes of an IP connection may not be the same, due to Network Address Translation.

Table 2: Requirements for sufficient observable information for acquisition and analysis for defence measures

Requirements	Notes
Middleboxes for endpoint defence should be able to determine a meaningful identifier for each endpoint and session.	The ability to distinguish sessions in a meaningful manner is critical to detect compromised endpoints and cyber attack.
Endpoints may have quality of service expectations of the middleboxes.	Where these performance expectations are negotiated, middleboxes should be able to estimate their own performance to indicate whether they are achievable.
Middleboxes for endpoint defence should be able to carry out the required analysis on transiting traffic without adding unreasonable latency. What is considered reasonable will be protocol/content dependent.	The level of defensive analysis/content validation, in order to meet these requirements, will be different for different sessions. Determining the context of each sessions reliably based on endpoints and protocols becomes highly important.

The current state of the art for instituting cyber defence measures in Clause 5.2 requires the following essential observable information at gateway middleboxes for acquisition and analysis:

- Identifiers and class/device type identifiers for service discovery, level of trust, and ownership attestation.
- Tuples defining IP addresses and ports.

- Domain names.
- URLs - both full and wildcarded to protect private queries.
- Application-related information such as Application Name, Process Name and ID, Username, Hashes of application-related files and libraries and application type.
- Hashes of files sent on network.
- Client information such as names and versions of Hardware, OS, Software and device type.
- Logging references.
- Accurate observable time stamps.

5.4 Ability to institute defence measures as part of gateway management

The set of observables obtained will facilitate cyber security decisions at the middlebox in response to policy configuration. The middlebox is likely to be but one part of the defensive mechanism of an enterprise, and it is expected to therefore be communicating with other security elements. This is out of scope here but noted as a richness of the middlebox architecture that may be useful as part of an implementation means for security attestation and ownership authentication.

For cyber security reasons the middlebox may wish to take action at any point, and any mechanisms should support this: e.g. stop connection, record sessions, alert administrator, warn/alert user, change security polices/configurations as a result. Middlebox security protocol aware endpoint devices should be capable of handling any of the potential actions that an MSP-conformant middlebox takes with no reduction in endpoint security. Middleboxes should not require a weakening of any cybersecurity protective measures on MSP-conformant endpoint devices for their operation (see TLS proxy that breaks EV certificate validation).

6 New challenges and mechanisms for gateway cyber defence

6.1 Introduction

This clause describes the new challenges being faced for gateway cyber defence and provides a survey of new techniques to address them. This knowledge helps in developing subsequent recommendations.

6.2 Challenges

6.2.1 Virtualization implementations

One of the largest scale and most far reaching transitions underway today is the transformation of network infrastructures globally to virtualized equivalents operated from data centres. This transition is being grounded on specifications developed by more than a dozen industry standards bodies known as Network Functions Virtualisation - Software Defined Networks. NFV is the equivalent of legacy signalling systems, while SDNs are the equivalent of legacy physical transmission and switch facilities. One of the principal collaborating venues is ETSI's Network Functions Virtualisation Industry Standards Group. See https://portal.etsi.org/tb.aspx?tbid=789&SubTB=789.

As discussed further in Clause 6.3, there are additional threat vectors for hypervisors and virtualization implementations. Virtual gateways are privileged, and if it is co-hosted require special separation and trust platform techniques. This architecture could also give rise to additional vulnerabilities and require separation from other VMs and VNFs. One of the remedies for secure middlebox implementations in a virtual environment is as described by the mbTLS initiative, below.

Virtual gateway middleboxes have a role in cyber security of the VMs/VNFs, i.e. authorizing running code and NFV-SDN configurations. The requirements and a model were recently adopted in ITU. See e.g. Recommendation ITU-T X.1038 [i.14].

6.2.2 5G mobile systems

The implementation and ubiquitous use of 5G with its reliance on NFV and Mobile Edge Computing will bring critical infrastructures onto commercial networks and give rise to enhanced cyber security requirements. Considerable work is occurring in multiple bodies, especially the ETSI NFV ISG and 3GPP SA5. See especially EU, 5G PPP Architecture Working Group, View on 5G Architecture [i.15]. The 5GPP Final Report includes multiple requirements and architectural implementations and techniques relating to virtualized defence gateways.

"One specific class of network functions in 5G will be the "Virtual Network Functions (VNFs)". They are represented by one or more virtual machines running different software and processes on top of industry-standard high-volume compute platforms, switches and storage, or cloud computing infrastructure. These are capable of implementing network functions traditionally implemented via custom hardware appliances and middleboxes (e.g. router, NAT, firewall, load balancer, etc.). The VNFs will play an important role especially in the design of CN functions." § 4.1, 5GPPP Final Report.

The Report's table 3, Potential service-specific flavours of network functions, enumerates core network functions for a broad array of gateway middlebox capabilities.

Table 3: Potential 5G service-specific flavours of network functions

Core network	Value added services	Parental Control (e.g., user context for children and requested service dependent optional part of a service chain), DPI, Video optimizers, firewalls, service chaining in GiLAN.
	Authentication, Authorization, Accounting (AAA) and security	Service-specific access control and accounting/charging policy functionality and placement. Service specific security (e.g. a slice with no encryption and/or with added data integrity).
	Traffic control	QoE, QoS, mapping, monitoring, flow processing/policing and enforcement done in a service-specific way. Tighter mobile network – Transport interaction.
	Mobility management	Mobility management function design and selection may be service-specific, to allow for a higher degree of customization, e.g., network-slice-specific or radio-access-specific mobility management

6.2.3 Autonomous Internet of Things (IoT) deployments

Lifecycle management of autonomous devices is an additional challenge, and one in which gateways have a significant cyber security role to play. One of the most extensive IoT attacks on global network infrastructure occurred on 21 October 2016 by an alleged 150 000 IoT devices that generated a one Terabit per second attack on multiple network devices. See http://securityaffairs.co/wordpress/51726/cyber-crime/ovh-hit-botnet-iot.html.

Known vulnerability of IoT devices and their growing threat has led to an array of activities in academic communities and standards bodies aimed at limiting the potential threats using middlebox gateway implementations. See, e.g. Proceedings of the 2016 ACM Workshop on Cloud-Assisted Networking (CAN'16) on Cloud-Assisted Networking [i.21], pages 55-60, Hafeez et al., Securebox: Toward Safer and Smarter IoT Networks [i.22], Stanford Annual Meeting : 2016 IoT Workshop, http://forum.stanford.edu/events/2016iot_workshop.php, especially Auditing IoT Communications with TLS-RaR, https://forum.stanford.edu/events/2016/slides/iot/Judson.pdf, and Embark: Securely Outsourcing Middleboxes to the Cloud [i.23].

In ITU-T SG20 treating IoT security, over an 18 month period, there were nearly 250 contributed documents treating IoT gateways, and many specifications for gateway based security are in progress. See, e.g. draft Recommendation ITU-T Y.gw-IoT-arch [i.24], draft Recommendation ITU-T Y.IoT-cdn [i.25]. Even greater numbers of work items are underway in 3GPP and oneM2M. Across all of these initiatives and bodies, middlebox gateway based security is viewed as essential in the face of the ever increasing IoT devices.

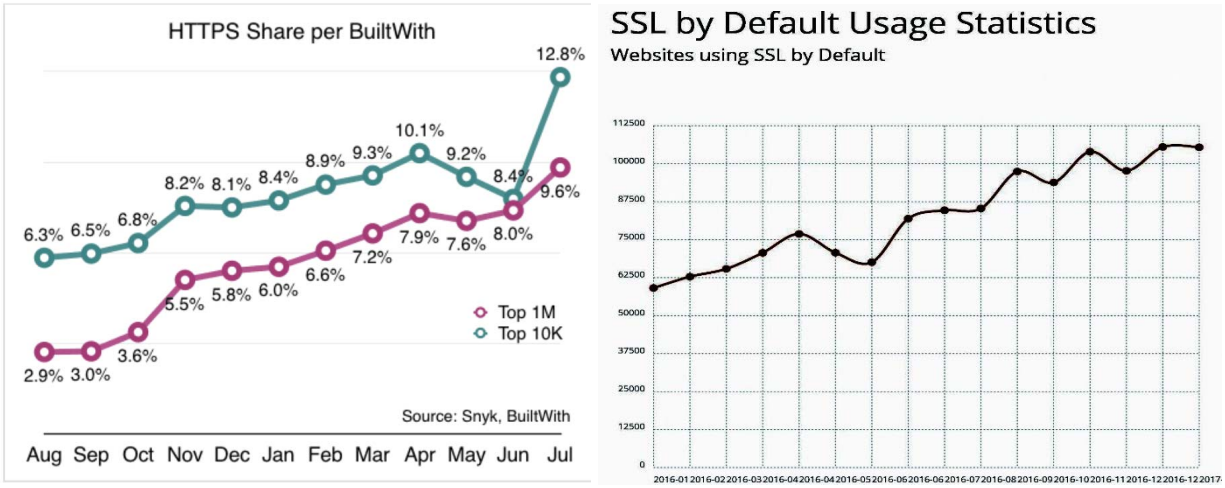
6.2.4 Over The Top (OTT) services

Policy doctrines such as NetNeutrality have resulted in the proliferation of Over The Top service providers. See ResearchGate, Impact of Over the Top (OTT) Services on Telecom Service Providers [i.26]. Because such services are essentially unregulated today in most national jurisdictions and the investment in infrastructure is minimal, there are significant incentives to use encryption as a means of differentiation and attracting customers as well as reinforcing bandwidth non-discrimination by the underlying transport carriers. OTT providers may be additionally incentivised to use encryption for providing services globally that would not be allowed in some jurisdictions - at least not on an unregulated basis. The market is divided into voice/conferencing providers, text/eMail providers, games, and mass media music and video providers. The Over The Top (OTT) market is estimated to grow from USD 28,04 Billion in 2015 to USD 62,03 Billion by 2020 with a CAGR of 17,2 %. The report on the OTT market considered 2014 as the base year and the forecast period from 2015 to 2020 [i.27].

As a result of these dynamics, the adverse effects suffered by underlying carriers surfaced at the IAB MaRNEW workshop described in more detail in clause 6.2.5. Especially relevant was the result of "no participation methods of OTTs in encryption management at middleboxes". See [i.28], section 5.1.

6.2.5 Widespread use of TLS as part of "Encrypt Everything" initiatives

As described in clause 6.3.1, an "encrypt everything" movement emerged over the past two years leading to increasing use of Transport Layer Security (TLS) between end user clients and data centre servers. Although HTTPS which makes use of TLS has been around for 20 years, it has always remained very lightly adopted until mid 2015. Data from reliable sources show HTTPS adoption has gone from 3 % to more than 10 % in a single year. The BuiltWith® web site metrics portal provides "SSL By Default" metrics that redirect visitors to HTTPS. Figure 2 portrays the increase - which continues into 2016. See BuiltWith®, SSL by Default Usage Statistics [i.29].



NOTE: Source: BuiltWith®

Figure 2: Adoption in top 1M sites 2015-2016, 2016

The adverse effects of this trend was already recognized by the Internet Architecture Board which called a workshop in September 2015 - Managing Radio Networks in an Encrypted World (MaRNEW) Workshop 2015 held together with the GSMA [i.30]. The IETF's own MARCOM workshop recognized that ubiquitous encryption has significant adverse effects on operating communication networks and providing services [i.4].

Since the MaRNEW workshop, encryption use for web based service has about doubled. As a result, BlueCoat-Fraunhofer findings have been underscored - that *"almost all modern security solutions such as intrusion detection systems (IDS), Intrusion Prevention Systems (IPS), web filters, Data Loss Prevention (DLP) systems or Next Generation Firewalls (NGFW) are reliant on being able to analyse unencrypted data communications. As a result, cyber criminals and malware creators increasingly use encryption to disguise their data communications, complicate analyses and bypass security systems unnoticed using compromised computers."* See BlueCoat & Fraunhofer Institute for Communication, Information Processing and Ergonomics FKIE, White Paper on Encrypted Traffic Management, January 2016 [i.31].

6.3 New and modified middlebox security protocol techniques

6.3.1 Introduction

This clause provides a survey of the extensive industry and academic research and development efforts that have produced a combination of existing protocol adaptations and effective new protocols and techniques to deal with the challenges in the clauses, above. One of these techniques - known as mcTLS - is relatively mature and has significant support as a middlebox security protocol, including their use for gateway defence.

The increasing use of Transport Layer Security (TLS) between end user clients and data centre servers has resulted in the widespread deployment of middleboxes that break TLS to counter the considerable adverse effects of the protocol's use. The technique is portrayed in Figure 3. Ref, David Naylor: Balancing Privacy and Functionality - Secure Communication with Middleboxes. cmuCyLab, https://www.youtube.com/watch?v=1YbztPssYk4&feature=youtu.be.

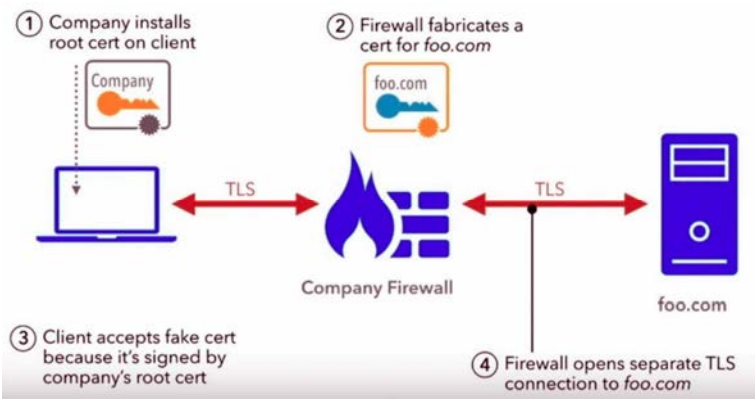


Figure 3: Example of Man-In-The-Middle (MITM) technique to defeat TLS

Finding a balance between MITM techniques and the need for effective cyber defence at gateways among many other essential requirements for traffic monitoring and shaping have led to the considerable work described below.

6.3.2 Multi-Context Transport Layer Security (mcTLS)

Middleboxes have become ubiquitous in today's telecommunication networks for a wide variety of essential needs described in Clause 4. The deployment of TLS discussed in Clause 6.2.5, however, has seriously impacted the use of middleboxes. Furthermore, TLS is only superficially secure as it was created for use between two parties and has no mechanism to authenticate middleboxes, nor do user clients have any security guarantees past middleboxes. Middleboxes have full read/write access for traffic passing through them.

The development of the Multi-Context Transport Layer Security (mcTLS) protocol was undertaken and funded as a cooperative research and development effort by prominent industry and academic institutions in both Europe and the United States [i.1]. The protocol builds on top of TLS to allow endpoints to explicitly and securely include in-network functionality to interact with middleboxes with complete visibility and control. mcTLS maintains TLS properties of providing:

1) entity authentication;

2) payload secrecy;

3) payload integrity;

while adding two more properties:

4) visibility & control;

5) least privilege.

The "least privilege" idea relies on the middlebox requirements being context and traffic dependent - leveraging the reality that most middleboxes do not need read/write access to all data. See Figure 4.

	Request		Response	
	Headers	*Body*	*Headers*	*Body*
Cache	○		●	●
Compression			●	●
Load Balancer	○			
IDS	○	○	○	○
Parental Filter	○			
Tracker Blocker	●		●	
Packet Pacer			○	
WAN Optimizer	○	○	○	○

(● = read/write; ○ = read-only)

NOTE: No middlebox needs read/write access to all of the data.

Figure 4: Examples of app-layer middleboxes and the permissions they need for HTTP (credits to David Naylor)

The visibility property is met by the inability of mcTLS to support transparent middleboxes - that is, those that are not visible to user client or server. Both end points need be aware of the middleboxes and the middleboxes ONLY receive a key for a context if both ends agree. The elegant flexibility of mcTLS' ability to enable interaction with middleboxes and contexts is illustrated in Figure 5. Ref, David Naylor: Balancing Privacy and Functionality - Secure Communication with Middleboxes. cmuCyLab, https://www.youtube.com/watch?v=1YbztPssYk4&feature=youtu.be.

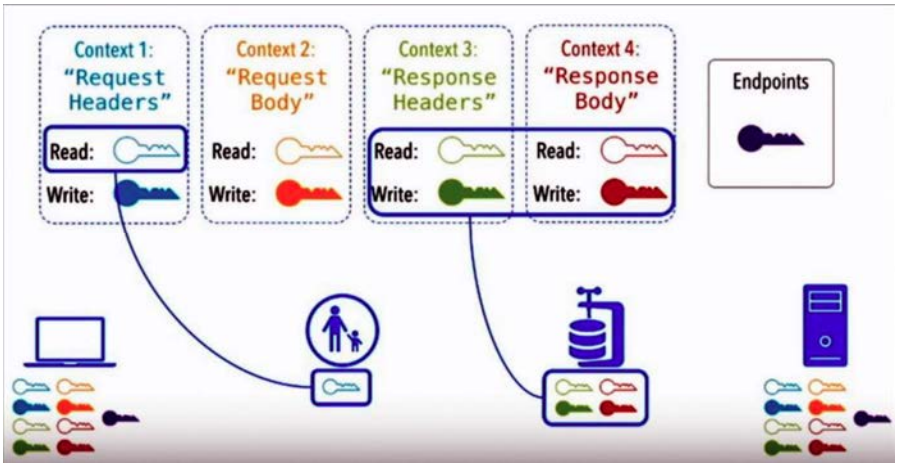


Figure 5: mcTLS middlebox access and encryption contexts for different read/write access control.

The performance impact and overhead for mcTLS are small. In addition, mcTLS does not require clients to have public keys - the same as TLS. Clients do not need authentication. If it is necessary, it can be done at the application layer. The end points both decide if they want a middlebox in the path. See also [i.32].

A variant of mcTLS tentatively referred to as mbTLS is also under active development. It is able to support legacy clients and does not require both endpoints be in agreement. mbTLS trades off joint control to get incremental deployability and is optimized for NFV-SDN datacentre based middlebox implementations. It supports two new deployment models: Middleboxes as a Service and Middleboxes at the Edge. It is attractive because of the significant economy of scale because each network does not need to buy their own middleboxes.

The Middlebox as a Service model, however, means that software and hardware providers in a cloud data centre are no longer the same. It introduces a new party to the threat model. A cloud trusted computing platform such as SGX where running code exists in "enclaves" and both the code and data is encrypted before it goes to memory and enables remote attestation. The technique, however, requires splitting the virtual middleboxes - which adversely affects performance. Ref, David Naylor, cmuCyLab, above. See also [i.33].

6.3.3 Other new protocol and structured expression platforms for middlebox security

Application-layer protocol negotiation. Application-layer protocol negotiation enables the broadcast of all needed information during/after handshake without the need to change/extend existing TLS protocol. The approach relies on an ALPN (Application-Layer Protocol Negotiation). A registration process would be created for ALPN Protocol Identifiers. Communication would start with a normal TLS handshake (ETI uses a fake certificate) and as a second step the client will not close the connection (like today if it does not like the CA), but sends some magic string to the ETI and expects a magic answer. If this answer is missing, ETI is not able to understand the protocol and client will disconnect. See Apple® Developer Conference, What's New in Security, https://developer.apple.com/videos/play/wwdc2016/706/?time=482.

Application Transparent Transport Layer Security (AT-TLS). Application Transparent Transport Layer Security (AT-TLS) creates a secure session on behalf of an application. Instead of implementing TLS in every application that requires a secure connection, AT-TLS provides encryption and decryption of data based on policy statements that are coded in the Policy Agent. The application sends and receives cleartext (unencrypted data) as usual while AT-TLS encrypts and decrypts data at the TCP transport layer. For more information about AT-TLS and AT-TLS policy setup, see the Application Transparent Transport Layer Security (AT-TLS). See http://www.ibm.com/support/knowledgecenter/en//SSLTBW_2.1.0/com.ibm.zos.v2r1.halx001/transtls.htm.

Blindbox. Blind-Box performs deep-packet inspection directly on the encrypted traffic. BlindBox realizes this approach through a new protocol and new encryption schemes. BlindBox enables applications such as IDS, exfiltration detection and parental filtering, and supports real rulesets from both open-source and industrial DPI systems. BlindBox has been shown to be practical for settings with long-lived HTTPS connections. Moreover, its core encryption scheme is 3-6 orders of magnitude faster than existing relevant cryptographic schemes. https://eprint.iacr.org/2015/264.pdf.

Cyber Threat Intelligence Cyber Observables (CybOX). The Cyber Observable Expression (CybOX) is a standardized language for encoding and communicating high-fidelity information about cyber observables, whether dynamic events or stateful measures that are observable in the operational cyber domain. By specifying a common structured schematic mechanism for these cyber observables, the intent is to enable the potential for detailed automatable sharing, mapping, detection and analysis heuristics. The referenced material serves as an overview of those specifications and defines how they are used within the broader CybOX framework. See https://github.com/CybOXProject/specifications/wiki. CybOX also leverages considerable related open source techniques and implementations associated with Structured Threat Information eXchange (STIX) and Malware Attribute Enumeration and Characterization (MAEC™). See https://wiki.oasis-open.org/cti/Open%20Source%20Projects and https://github.com/MAECProject/.

Device chain inspection. Device chain inspection enables each device in a chain of network devices to get information about all devices, i.e. all inspecting devices should be visible. The approach can be met without TLS protocol modifications by invoking two different levels of trust. One can either trust the next hop inspecting device (identified by its certificate chain) and get all data about other devices forwarded by it. Alternatively, a proof can be required that the data forwarded by these devices is correct, e.g. using an iterative process and some point-to-point communication from client to each ETI device, so that the client can verify the data given by its neighbour. See https://ieee-sa.imeetcentral.com/etiwg/folder/WzIwLDY5NTI5NTZd/WzIsNDY2NDE1MThd/.

Explicit Trusted Proxy. Explicit Trusted Proxy provides two alternative methods for a user-agent to automatically discover and for a user to provide consent for a Trusted Proxy to be securely involved when he or she is requesting an HTTP URI resource over HTTP2 with TLS. The consent should be per network access. The technique also includes the role of the Trusted Proxy such as the access service provider in helping the user to fetch HTTP URIs resource when the user has provided consent to the Trusted Proxy to be involved. See https://tools.ietf.org/html/draft-loreto-httpbis-trusted-proxy20-01.

FlowTags. FlowTags provide a flow tracking capability to ensure consistent policy enforcement in the presence of middlebox dynamic traffic modifications. FlowTags are an extended SDN architecture in which middleboxes add Tags to outgoing packets, to provide the necessary causal context (e.g., source hosts or internal cache/miss state). These Tags are used on switches and (other) middleboxes for systematic policy enforcement. Minimally extending middleboxes can provide this support. Challenges include the design of southbound and northbound FlowTags APIs, new controllayer applications for enforcing and verifying policies, and automatically modifying legacy middleboxes to support FlowTags. See http://www.cs.columbia.edu/~lierranli/coms6998-10SDNFall2014/papers/Flowtags-HotSDN2013.pdf.

Framework for Consent and Permissions. This technique attempts to provide a kind of meta framework for several other techniques through classification into five different types: Preconfigured Client Consent; Dynamic Client Consent; Preconfigured Service Consent; Partial Read-only Consent, and Partial Read/Write Consent. See https://www.iab.org/wp-content/IAB-uploads/2015/08/MaRNEW_1_paper_31.pdf.

Integrated Adaptive Cyber Defence fabrics (IACD) and Open C2. Integrated Adaptive Cyber Defence (IACD) is an architecture which combined with the Open C2 command and control language, is intended to improve gateway cyber defence by enabling rapid detection and mitigation of cyber threats for perimeter boundary protection. IACD OpenC2 based defence implementations support automated, multi-part actions at machine speed sharing of indicators, coordination of responses between domains, synchronization of cyber defence mechanisms. See http://openc2.org/docs/OpenC2%20Effort%20Overview%20Base%2016%20Feb%202016.pdf.

Keyless SSL. Service side consents using Keyless SSL service gives the middlebox read and write access to traffic. Presence of the middlebox in the security association is transparent to the client. The TLS handshake is split geographically, with most of the handshake happening at the datacentre server edge while moving the private key operations to a remote key server. This key server can be put on the customer's infrastructure, giving them exclusive access to the private key. The key server only allows connections from clients with a certificate signed by a datacentre server provider's internal certificate authority. The certificates are granted by the provider's own certificate authority for both sides of this connection. The provider has strict controls over how these certificates are granted and use the X.509 Extended Key Usage option to ensure that certificates are only used as intended. This prevents any party that does not have a provider granted certificate from communicating with the key server. Customers also have the option to add firewall rules to limit incoming connections to those from provider's IP space. See https://blog.cloudflare.com/keyless-ssl-the-nitty-gritty-technical-details/.

MAMI; Middlebox Cooperation Protocol. Measurement and Architecture for a Middleboxed Internet (MAMI) is a European Commission Horizon 2020 funded research project to allow explicit cooperation between endpoints and middleboxes enabling appropriate in-network services to ease management and scalability. It is pursuing a Middlebox Cooperation Protocol that provides explicit cooperation (endpoint and application signalling to middleboxes and vice-versa). See https://mami-project.eu/index.php/2015/09/22/welcome-to-mami/. This work is an extension of IAB Workshop on Stack Evolution in a Middlebox Internet (SEMI) 2015. See https://www.iab.org/activities/workshops/semi/.

Middlebox-cooperative TCP for a non end-to-end internet (HICCUPS). In-path middleboxes that modify packet headers are typically transparent to TCP, yet can impact end-to-end performance or cause blackholes. TCP HICCUPS reveals packet header manipulation to both endpoints of a TCP connection. Handshake based Integrity Check of Critical Underlying Protocol Semantics (HICCUPS) permits endpoints to cooperate with currently opaque middleboxes without prior knowledge of their behaviour. For example, with visibility into end-to-end behaviour, a TCP can selectively enable or disable performance enhancing options. This cooperation enables protocol innovation by allowing new IP or TCP functionality (e.g. ECN, SACK, Multipath TCP, Tcpcrypt) to be deployed without fear of such functionality being misconstrued, modified, or blocked along a path. HICCUPS is incrementally deployable and introduces no new options. TCP HICCUPS was implemented across thousands of disparate Internet paths to show the breadth and scope of subtle and hard to detect middlebox behaviours encountered. Path diagnostic capabilities provided by HICCUPS can benefit applications and the network. See http://rbeverly.net/research/papers/hiccups-sigcomm14.pdf.

Mobile Throughput Guidance Inband Signalling Protocol. This technique consists of mechanism and protocol elements that allow the cellular network to provide near real-time information on capacity available to the TCP server. This "Throughput Guidance" (TG) information would indicate the throughput estimated to be available at the radio downlink interface (between the Radio Access Network (RAN) and the mobile device (UE)). TCP server can use this TG information to ensure high network utilization and high service delivery performance. The technique is also applicable for video delivery over cellular networks, and has been tested in live operations. See https://tools.ietf.org/html/draft-flinck-mobile-throughput-guidance-02.

NIMBLE. NIMBLE is optimized for SDN and allows network operators to specify a logical view of the middlebox policy and automatically translates this into forwarding rules that take into account the physical topology, switch capacities, and middlebox resource constraints. There are three key ideas underlying NIMBLE's design. First, efficient data plane support for composition, including tunnels between switches and using SDN capabilities to add tags to packet headers that annotate each packet with its processing state. Second, practical unified resource management, where the intractability of optimization is addressed by decomposing the problem into a hard offline component that accounts for the integer constraints introduced by switch capacities and an efficient online component that balances middlebox load in response to traffic changes. Third, learning middlebox dynamics where reporting capabilities of SDN switches is exploited to design lightweight flow correlation mechanisms that account for most common middlebox-induced packet transformations. See https://nsl.cs.usc.edu/~rmiao/publications/Zafar13_.pdf.

Out of band session key sharing. This technique specifies a method for connecting to a proxy via a secure channel, allowing, disallowing, and detecting any transforms that the proxy may perform, and allowing the proxy to connect via secure channel to another site on the user's behalf. See https://tools.ietf.org/html/draft-rpeon-httpbis-exproxy-00.

Substrate Protocol for User Datagrams (SPUD). Spud is a technique for selective information exposure designed to support transport evolution. Since the widespread deployment of Network Address Translation (NAT) has in effect moved the network-transport layer header boundary to the other side of the source and destination port numbers, and Transmission Control Protocol (TCP) brings a complex set of semantics over which new transports may not necessarily be mappable, demand that new transport protocol deployment occur over User Datagram Protocol (UDP). SPUD is realized as a shim between UDP and an (encrypted) transport protocol, and provides minimal sub-transport functionality, of use to devices on path by:

1) grouping of packets together into tubes;

2) signalling of the start and end of a tube, assisting in state setup and teardown along the path; and

3) an extensible signalling mechanism based on typevalue encoding in the Concise Binary Object Representation (CBOR) for associating properties with individual packets or all packets in a tube. See http://www.tik.ee.ethz.ch/file/84f1fc268860f31a4ed7e0345231b855/spud.pdf.

Transport Layer Security Auxiliary Data (TLS-AUX). This technique introduces a new layer between TCP and TLS, called TLS-AUX. By virtue of this position in the protocol stack, the TLS-AUX layer can recognize the data flowing on TLS connections in either direction as a sequence of structured TLS records. A new type of TLS record is introduced, called the AUX record, which is allowed to exist only at the TLS-AUX layer and not higher. The AUX records are to be used for communication of metadata between network services and webservers. Network services and webservers may directly insert and retrieve their metadata into the TLS connections in the form of AUX records, and retrieve them from the connections' flow. In this manner, the mechanism is very similar to network services inserting and removing HTTP request and response headers directly into the body of TCP connections. See https://www.akamai.com/us/en/multimedia/documents/technical-publication/tls-aux-associating-auxiliary-data-tls-connections.pdf.

Transport Layer Security Rotate and Release (TLS-RaR). TLS-RaR is a method for passive, read only auditing of TLS-protected communication, to replace the man in the middle method, and remove the TLS barrier from a communications audit. The technique is optimized for IoT. A device simply distributes key to Audit Boxes. Cryptographic Hashes and Signatures ensure integrity to the auditors. After a clean key exchange, the IoT application ensures the last key encrypting data is not useful (e.g. authenticated acknowledgment), then securely releases the key.

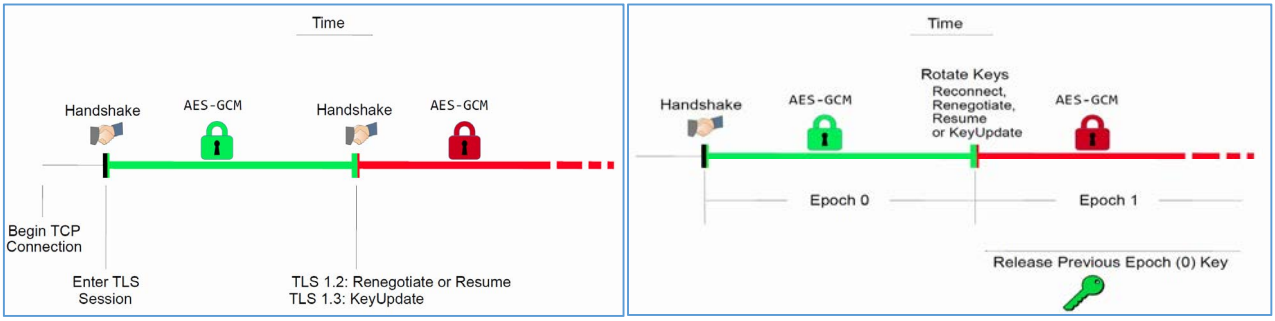


Figure 6

See https://forum.stanford.edu/events/2016/slides/iot/Judson.pdf.

TLS Proxy Server Extension. This technique defines a TLS extension that allows a TLS server to provide a TLS client with all of information about the other TLS server (or servers) that are participating in the application layer traffic that the client needs to make a well-informed access control decision. This empowers the client to reject TLS sessions that include servers that it does not trust. See https://tools.ietf.org/html/draft-mcgrew-tls-proxy-server-01.

7 Recommendations

7.1 Introduction

Drawing on the industry segment security requirements, the cyber industry best practice, and the interoperability initiatives already under way as described above, this clause sets forth recommendations to establish WHAT information is beneficial for cyber defence, WHERE it is known and WHERE it needs to be known. Recommendations are also provided for HOW it could be got there, by referring to how one could use existing protocols or initiatives, or by proposing new requirements and possible solutions to meet these gaps that can ensue, including subsequent ETSI TC CYBER work on Technical Specifications, profiles, and guides in collaboration with other bodies.

The recommendations do not include anything related to how the analysis is done on this data and how it is shared and acted upon across the ecosystem of cyber defence components. Active initiatives such as OASIS Cyber Threat Information define such sharing platforms.

7.2 Control at the gateway

Information that supports gateway cyber defence includes a description of WHAT is important, and WHERE this data is known - which could be multiple places, but in many cases will be at a client or server end point inside an encryption layer. Although some of this data is sensitive and may affect privacy, there is also an important cyber defence role in the availability of this data so an appropriate balance can be struck between cyber security and privacy. Any exposure needs be secure, and should allow the data owner to determine how it is treated.

Based on the use cases, derivative requirements, and survey of available techniques, WHAT is important in a middlebox security protocol may include the following capabilities:

Middlebox Security Protocol capability profile

- Ability of the client and server at a sufficient level of granularity to:
  - discover and identify all gateway middlebox devices that are potentially able to read/modify the traffic;
  - decide whether each middlebox can read and/or modify content;
  - detect any changes to the content;
  - detect who has done modifications;
  - obtain proof that it can see all devices without depending on the first middlebox;
  - configure different access rights for different blocks of traffic, e.g. header/body or different streams in HTTP2.
- Ability of the client to control whether the server can see beyond middleboxes (for client anonymity).
- Ability of an intercepting device to read traffic without the need for re-encryption.
- Ability of a client to support a post mortem analysis.
- Transparency for load balancers without requiring out-of-band communication.
- Acceptable overhead (performance, usability).
- A sufficient degree of backward compatibility with different clients.

7.3 Observable availability at the gateway

WHAT is important in a middlebox security protocol also includes the ability to obtain the following observables at the gateway with a sufficient level of granularity:

Observables

- Identifiers and class/device type identifiers for service discovery, level of trust, and ownership attestation.
- [end point identifier] tuples for IP addresses and ports and mappings.
- Domain names.
- URLs - both full and wildcarded to protect private queries.
- Application-related information such as Application Name, Process Name and ID, Username, Hashes of application-related files and libraries, application type.
- Hashes of files sent on network.
- Client information such as names and versions of Hardware, OS, Software, device type.
- Logging references.
- Accurate observable time stamps.

HOW the observables get between the endpoint and gateway is very important, both practically because this is likely to need a newly defined network channel, but also from a security point of view because this connection needs to meet the technical requirements of Clause 5.

The recommended roadmap for HOW is in two phases.

Phase 1 - new out of band channel

In the absence of an existing protocol serving the needs of observable transport, a new secure channel between the end point and the gateway is recommended to be implemented out of band from the data sessions.

The channel may be used to send different protocols depending on the observable. For some observables, there are protocols which already offer suitability and could be bought close to meeting the requirements.

For example, SIMEF helps transfer NAT mappings and can be used for greater detail such as URLs and hardware/software attestation/reporting. New schema may be required (e.g. in XML/JSON) but in the first instance one can try to reuse others. See Recommendation ITU-T X.1542 [i.34].

OASIS STIX offers protocols for observables. As does OpenC2 and Integrated Adaptive Cyber Defence fabrics (IACD). ITU-T CYBEX also offers protocols for observables. Any of these might be usefully considered to define the protocol format of observables on the channel.

The security requirements for information access to the channel would need to be specific: how should authentication of the gateway and endpoint work, and what granular control should the owner have over the observables passed to the gateway.

A new out of band channel has limitations (observables are asynchronous to session data, observables do not necessary following same path through middleboxes as the session data) and benefits (there are fewer dependencies on existing channel specifications so the new channel is much quicker to implement in a manner that meets all the privacy and security requirements).

Of critical importance to an out of band channel is an identification means to associate observables on the out of band channel with their in-band data sessions.

Phase 2 - adapt existing in-band channels

This phase is the next level of uptake of flows of observables between endpoints and gateways, where it becomes more seamlessly integrated with existing protocol stacks. As such, this second phase would see deployment of channels that are in-band with the session data in order to overcome the principle drawback of an out of band channel by assuring that the observable is always delivered to the gateway alongside its session data and at the very same time.

The channels and protocols specified in this second phase would be more mature and would place relevant cyber observables into the protocols of the session itself - for example, carrying the securely protected observable in a header or extension packet. The advantage of keeping the cyber observable information in-line as part of the session it is describing, is that association to the session and access to the observable are guaranteed by authorized middleboxes. Interoperability and adoption is easier because one does not need the endpoint to support an additional new and separate protocol and connection.

Existing protocols with extensible headers could be used where appropriate, and the choice will depend on the observable. However promising options are TCP Mobile Throughput Guidance channel which is a slow secure channel for use between network components, and the TLS -AUX channel which exists between TCP and TLS.

7.4 Adoption of a common Middlebox Security Protocol, profiles and guidelines

There is plainly a rapidly growing, urgent need for a stable, industry supported middlebox security protocol for gateway cyber defence. The extensive survey and analysis undertaken in the present document suggest that the mcTLS protocol is a good initial candidate for meeting that need. At the same time, it is also apparent that other middlebox security protocols are emerging - especially for NFV-SDN implementations. While there are many possible venues in which mcTLS could emerge as a published industry specification, the benefits of ETSI itself engaging in this activity could significantly facilitate the speed and success of its adoption. It is therefore recommended that TC CYBER undertake this activity as a multi-part specification that allows for evolution of the protocols and development of additional techniques.

As part of this continuing effort, profiles and guidelines should also be included as part of the work. It may also be necessary, as described in Clause 7.3, to identify suitable means for the HOW dimension of accessing the information and transferring it in a controlled and trusted manner to where it needs to be delivered.

7.5 Specification of a new out-of-band secure channel between endpoint and gateway, and protocols for a set of observables

There is a need for a secure communication channel to be specified between endpoint and gateway. There are major initiatives on cyber observable sharing between cyber security entities, with advanced and complicated specifications, but there is a gap for something deployable more simply, compactly and readily, on the range of widespread endpoint platforms. This new specification would get observables from the endpoint to the gateway, and hereby catalyse the remainder of cyber security information sharing and event detection between cyber security entities.

An out-of-band channel would be specified, with security controls that meet the technical requirements of Clause 5 so that the endpoint and gateway are in control of what they choose to share. Also specified would be a limited set of observables, and protocol formats to transmit these observables by the channel. STIX is a promising source from which to select a minimal subset of observables and their respective protocol formats. A means to associate observables with their data session would also be specified, this would be a description so the gateway could identify the associated session. The deliverables are set out as follows:

Deliverables

- Secure authorized channel between endpoint and gateway.
- Set of observables and protocol specifications.
- Means to associate observables to data session.

The aim is rapid time to adoption of a secure means to communicate a minimal but authorized and useful set of observables between endpoint and gateway; this will have greatest immediate benefit to enterprise cyber security by standardizing a suitable interoperable means for cyber observables to be passed from enterprise devices to enterprise middleboxes.

Note, while mcTLS might act at the application layer to permit read or write access to subsets of plaintext context, this channel would allow additional cyber observables from different levels of privilege, such as application-related information (application name, process name and ID, username, file and library hashes, application type), and device class types and identifiers, and client information (hardware, OS, software, device type), and logging references.

7.6 Encouraging use of gateway cyber defence capabilities

ETSI should additionally use multiple available means for encouraging use of the protocol. Such means would include collaboration with the many industry, academic, and public policy communities focusing on middlebox security requirements, challenges and techniques. The technique should also be included as part of the suite of specifications necessary to the implementation of the EU Network and Information Security Directive [i.2].

Annex A: Bibliography

NOTE: This bibliography is significantly imported from reference [i.1]

- ACM CoNEXT '09, pages 1-12, New York, NY, USA: "Networking named content".
- ACM CoNEXT '14, pages 133-140, New York, NY, USA, 2014:"The Cost of the "S" in HTTPS".
- ACM Hotnets-IX, pages 9:1-9:6, New York, NY, USA, 2010: "Using strongly typed networking to architect for tussle".
- ACM MobiSys '10, New York, NY, USA, 2010.: "Catnap: exploiting high bandwidth wireless interfaces to save energy for mobile devices".
- ACM MobiSys '13, pages 319-332, New York, NY, USA, 2013: "Comparison of caching strategies in modern cellular backhaul networks".
- ACM MobiSys '14, pages 218-231, New York, NY, USA, 2014: "Characterizing resource usage for mobile web browsing".
- ACM SIGCOMM '12, pages 13-24, New York, NY, USA, 2012: "Making middleboxes someone else's problem: Network processing as a cloud service".
- ACM StorageSS '05, pages 51-56, New York, NY, USA, 2005: "Toward securing untrusted storage without public-key operations".
- ACM Trans. Comput. Syst., 2(4):277-288, Nov. 1984: "End-to-end arguments in system design".
- CCNC, pages 891-895. IEEE, 2011: "On the energy efficiency of proxy-based traffic shaping for mobile audio streaming".
- IEEE, Internet Computing, 15(2):27-34, March 2011: "To cache or not to cache: The 3g case".
- IEEE, Proceedings of the IEEE, 63(9):1278-1308, 1975: "The protection of information in computer systems".
- IETF RFC 5246 (as updated): "The Transport Layer Security (TLS) Protocol Version 1.2".
- IETF RFC 6762: "Multicast DNS".
- IETF RFC 6783: "DNS-Based Service Discovery".
- International Journal of Applied Cryptography, 2(2):154-158, 2010: "On reusing ephemeral keys in diffie-hellman key agreement protocols".
- NSDI '15, pages 367-380, Oakland, CA, May 2015. USENIX Association: "Flywheel: Google's data compression proxy for the mobile web".
- PAM '15: "Investigating transparent web proxies in cellular networks".
- SIGCOMM CCR, Jan. 2001: "Rethinking the tcp nagle algorithm".
- Springer, Passive and Active Measurement, 2014, pages 183-192: "Here be web proxies".
- USENIX Security'10, Berkeley, CA, USA, 2010. USENIX Association: "The case for ubiquitous transport-level encryption".
- UW-CSE-12-09-05, University of Washington, Sept. 2012, Technical Report: "An Internet Architecture Based on the Principle of Least Privilege".
- University College London PhD Thesis, 2001, I. Brown: "End-to-end security in active networks".

History

Document history		
V1.1.1	April 2017	Publication