Network Working Group
Internet-Draft
Intended status: Informational
Expires: September 11, 2017

K. Moriarty
Dell EMC
A. Morton
AT&T Labs

March 10, 2017

# Effect of Pervasive Encryption
# draft-mm-wg-effect-encrypt-08

## Abstract

Increased use of encryption impacts operations for security and network management causing a shift in how these functions are performed. In some cases, new methods to both monitor and protect data will evolve. In other cases, the ability to monitor and troubleshoot could be eliminated. This draft includes a collection of current security and network management functions that may be impacted by the shift to increased use of encryption. This draft does not attempt to solve these problems, but rather document the current state to assist in the development of alternate options to achieve the intended purpose of the documented practices.

## Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 11, 2017.

Table of Contents

## 1. Introduction

In response to pervasive monitoring revelations and the IETF consensus that Pervasive Monitoring is an Attack [RFC7258], efforts are underway to increase encryption of Internet traffic. Session encryption helps to prevent both passive and active attacks on transport protocols; more on pervasive monitoring can be found in the Confidentiality in the Face of Pervasive Surveillance: A Threat Model and Problem Statement [RFC7624]. The Internet Architecture Board (IAB) released a statement advocating for increased use of encryption in November 2014. Views on acceptable encryption have also shifted and are documented in "Opportunistic Security" (OS) [RFC7435], where cleartext sessions should be upgraded to unauthenticated session encryption, rather than no encryption. OS encourages upgrading from cleartext, but cannot require or guarantee such upgrades. Once OS is used, it allows for an upgrade to authenticated encryption. These efforts are necessary to improve end user's expectation of privacy, making pervasive monitoring cost prohibitive. Active attacks are still possible on sessions where unauthenticated sessions are in use. The push for ubiquitous encryption via OS is specific to improving privacy for everyday users of the Internet.

Although there is a push for OS, there is also work being done to improve implementation development and configuration flaws of TLS and DTLS sessions to prevent active attacks used to monitor or intercept session data. The (UTA) working group is in process of publishing documentation to improve the security of TLS and DTLS sessions. They have documented the known attack vectors in [RFC7457] and have documented Best Practices for TLS and DTLS in [RFC7525] and have other documents in the queue.

Estimates for session encryption from spring 2015 approximate that about 30% of web sites have session encryption enabled, according to

the Electronic Frontier Foundation [EFF]. The Mozilla Foundation maintains statistics on SSL/TLS usage and as of March 2015, 64% of HTTP transactions are encrypted. Enterprise networks such as EMC observe that about 78% of outbound employee traffic was encrypted in June 2014. Although the actual number of sites may only be around 30%, they include some of the most visited sites on the Internet for corporate users.

In addition to encrypted web site access (HTTP over TLS), there are other well-deployed application level transport encryption efforts such as mail transfer agent (MTA)-to-MTA session encryption transport for email (SMTP over TLS) and gateway-to-gateway for instant messaging (XMPP over TLS). Although this does provide protection from transport layer attacks, the servers could be a point of vulnerability if user-to-user encryption is not provided for these messaging protocols. User-to-user content encryption schemes, such as S/MIME and PGP for email and encryption (e.g. Off-the-Record (OTR)) for Extensible Messaging and Presence Protocol (XMPP) are used by those interested to protect their data as it crosses intermediary servers, preventing the vulnerability described by providing an end-to-end solution. User-to-user schemes are under review and additional options will emerge to ease the configuration requirements, making this type of option more accessible to non-technical users interested in protecting their privacy.

Increased use of encryption (either opportunistic or authenticated) will impact operations for security and network management, causing a shift in how these functions are performed. In some cases new methods to monitor and protect data will evolve, for other cases the need may be eliminated. This draft includes a collection of current security and network management functions that may be impacted by this shift to increased use of encryption. This draft does not attempt to solve these problems, but rather document the current state to assist in the development of alternate options to achieve the intended purpose of the documented practices.

In this document we consider several different forms of service providers, so we distinguish between them with adjectives. For example, network service providers (or network operators) provide IP-packet transport primarily, though they may bundle other services with packet transport. Alternatively, application service providers primarily offer systems that participate as an end-point in communications with the application user, and hosting service providers lease computing, storage, and communications systems in datacenters. In practice, many companies perform two or more service provider roles, but may be historically associated with one.

## 2. Network Service Provider Monitoring

Network Service Providers (SP) are responding to encryption on the Internet, some helping to increase the use of encryption and others preventing its use. Network SPs for this definition include the backbone Internet Service providers as well as those providing infrastructure at scale for core Internet use (hosted infrastructure and services such as email).

Following the Snowden revelations, application service providers responded by encrypting traffic between their data centers to prevent passive monitoring from taking place unbeknownst to the providers (Yahoo, Google, etc.). Large mail service providers also began to encrypt session transport to hosted mail services. This had an immediate impact to help protect the privacy of users data, but created a problem for network operators. They could no longer gain access to session streams resulting in actions by several to regain their operational practices that previously depended on cleartext data sessions.

The EFF reported [EFF2014] several network service providers taking steps to prevent the use of SMTP over TLS by breaking STARTTLS (section 3.2 of [RFC7525]), preventing the negotiation process resulting in fallback to the use of clear text. The use of encryption prevents middle boxes from performing functions that range from some methods of load balancing to monitoring for attacks or enabling "lawful intercept", such that described in [ETSI101331] and [CALEA] in the US. These practices are representative of the struggles administrators have with changes in their ability to monitor and manage traffic.

### 2.1. Middlebox Monitoring

Network service providers use various monitoring techniques for security and operational purposes. The following subsections detail the purpose of each type of monitoring and what protocol fields are used to accomplish the task.

#### 2.1.1. Load Balancers

Some network architectures need to share significant traffic load among a pool of parallel systems, to achieve the needed capacity. Load Balancer devices (a form of middlebox) provide the traffic-sharing function, according to pre-defined rules. A general rule for many load balancers requires that all packets comprising an individual flow should to be routed to the same system in the load balancer's pool. The definition of a flow will be based on a combination of header fields, often as many as five for 5-tuple flows

(including addresses and ports for source and destination, and one additional field such as the DSCP or other priority marking). Encryption that conceals or replaces the original IP header and/or transport header with modified addresses or ports may result in a set of flows being treated as one for load balancing purposes, which could cause uneven traffic load levels in the pool and unnecessary congestion when capacity limits are approached.

### 2.1.2. Traffic Analysis Fingerprinting

Fingerprinting is used in traffic analysis and monitoring to identify traffic streams that match certain patterns. This technique may be used with clear text or encrypted sessions. Some Distributed Denial of Service (DDoS) prevention techniques at the Network SP level rely on the ability to fingerprint traffic in order to mitigate the effect of this type of attack. Thus, fingerprinting may be an aspect of an attack or part of attack countermeasures.

The first/obvious trigger for DDoS mitigation is uncharacteristic traffic volume and/or congestion at various points associated with the attackee's communications. One approach to mitigate such an attack involves distinguishing attacker traffic from legitimate user traffic through analysis. The ability to examine layers and payloads above transport provides a new range of filtering opportunities at each layer in the clear. Fewer layers are in the clear means reduced filtering opportunities to mitigate attacks.

Traffic analysis fingerprinting could also be used on web traffic to perform passive monitoring and invade privacy.

For example, browser fingerprints are comprised of many characteristics, including User Agent, HTTP Accept headers, browser plug-in details, screen size and color details, system fonts and time zone. A monitoring system could easily identify a specific browser, and by correlating other information, identify a specific user.

### 2.1.3. Traffic Surveys

Internet traffic surveys are useful in many well-intentioned pursuits, such as CAIDA data [CAIDA] and SP network design and optimization. Tracking the trends in Internet traffic growth, from earlier peer-to-peer communication to the extensive adoption of unicast video streaming applications, has required a view of traffic composition and reports with acceptable accuracy. As application designers and network operators both continue to seek optimizations, the role of traffic surveys from passive monitoring grows in importance.

Passive monitoring makes inferences about observed traffic using the maximal information available, and is subject to inaccuracies stemming from incomplete sampling (of packets in a stream) or loss due to monitoring system overload. When encryption conceals more layers in each packet, reliance on pattern inferences and other heuristics grows, and accuracy suffers. For example, the traffic patterns between server and browser are dependent on browser supplier and version, even when the sessions use the same server application (e.g., web e-mail access). It remains to be seen whether more complex inferences can be mastered to produce the same monitoring accuracy.

### 2.1.4. Deep Packet Inspection (DPI)

The features and efficiency of some Internet services can be augmented through analysis of user flows and the applications they provide. For example, network caching of popular content at a location close to the requesting user can improve delivery efficiency (both in terms of lower request response times and reduced use of International Internet links when content is remotely located), and authorized parties use DPI in combination with content distribution networks to determine if they can intervene effectively. Web proxies are widely used [WebCache], and caching is supported by the recent update of "Hypertext Transfer Protocol (HTTP/1.1): Caching" in [RFC7234]. Encryption of packet contents at a given protocol layer usually makes DPI processing of that layer and higher layers impossible.

Data transfer capacity resources in cellular radio networks tend to be more constrained than in fixed networks. This is a result of variance in radio signal strength as a user moves around a cell, the rapid ingress and egress of connections as users hand-off between adjacent cells, and temporary congestion at a cell. Mobile networks alleviate this by queuing traffic according to its required bandwidth and acceptable latency: for example, a user is unlikely to notice a 20ms delay when receiving a simple Web page or email, or an instant message response, but will very likely notice a re-buffering pause in a video playback or a VoIP call de-jitter buffer. Ideally, the scheduler manages the queue so that each user has an acceptable experience as conditions vary, but the traffic type has been required to be known to date. Application and transport layer encryption make the traffic type detection less accurate, and affect queue management.

### 2.1.5. Connection to Proxy for Compression

In contrast to DPI, various applications exist to provide data compression in order to conserve the life of the user's mobile data plan and optimize delivery over the mobile link. The compression proxy access can be built into a specific user level application, such as a browser, or it can be available to all applications using a system level application. The primary method is for the mobile application to connect to a centralized server as a proxy, with the data channel between the client application and the server using compression to minimize bandwidth utilization. The effectiveness of such systems depends on the server having access to unencrypted data flows. As the percentage of connections using encryption increases, these data compression services will be rendered less effective, or worse, they will adopt undesirable security practices in order to gain access to the unencrypted data flows.

### 2.1.6. Mobility Middlebox Content Filtering

Service Providers may, from time to time, be requested by law enforcement agencies to block access to particular sites such as online betting and gambling, or access to dating sites. Content Filtering can also happen at the endpoints or at the edge of enterprise networks. This section is intended to merely document this current practice by operators and the effects of encryption on the practice.

Content filtering in the mobile network usually occurs in the core network. A proxy is installed which analyses the transport metadata of the content users are viewing and either filters content based on a blacklist of sites or based on the user's pre-defined profile (e.g. for age sensitive content). Although filtering can be done by many methods one common method occurs when a DNS lookup of a hostname in a URL which appears on a government or recognized block-list( [RFC7858] aims to address this). The subsequent requests to that domain will be re-routed to a proxy which checks whether the full URL matches a blocked URL on the list, and will return a 404 if a match is found. All other requests should complete.

See the Appendix for more information on "Encryption Impact on Mobility Network Optimizations and New Services".

#### 2.1.6.1. Parental Controls

Another form of content filtering is called parental control, where some users are deliberately denied access to age-sensitive content as a feature to the service subscriber. Some sites involve a mixture of universal and age-sensitive content and filtering software. In these

cases, more granular (application layer) metadata may be used to analyze and block traffic, which will not work on encrypted content.

#### 2.1.6.2. HTTP Redirection

There are cases (beyond parental control) when a mobile network service provider needs to redirect customer requests for content:

1. The mobile network service provider is performing the accounting and billing for the content provider, and the customer has not (yet) purchased the requested content.

2. Further contenty may not be allowed as the customer has reached their usage limit and needs to purchase additional data service.

Currently, the mobile network service provider redirects the customer using HTTP redirect to a page which educates the customer on the reason for the blockage and provide steps to proceed. Once the content is encrypted, the Mobile carrier loses the option to redirect the traffic leaving the option to block the customer's request and cause a bad customer experience untill the blocking reason can be conveyed by some other means. The customer may need to call customer care to find out the reason, both an inconvenience to the customer and additional overhead to the mobile network service provider.

### 2.1.7. Access and Policy Enforcement

#### 2.1.7.1. Server load balancing

Where network load balancers have been configured to route according to application-layer semantics, an encrypted payload is effectively invisible. This has resulted in practices of intercepting TLS in front of load balancers to regain that visibility, but at a cost to security and privacy.

#### 2.1.7.2. Network Access

Approved access to a network is a prerequisite to requests for Internet traffic - hence network access, including any authentication and authorization, is not impacted by encryption.

Cellular networks often sell tariffs that allow free-data access to certain sites, known as 'zero rating'. A session to visit such a site incurs no additional cost or data usage to the user. This feature may be impacted if encryption hides the details of the content domain from the network. This topic and related material are described further in the Appendix.

#### 2.1.7.3. Regulation and policy enforcement

Mobile networks (and usually ISPs) operate under the regulations of their licensing government authority. These regulations include Lawful Intercept, adherence to Codes of Practice on content filtering, and application of court order filters.

These functions are impacted by encryption, typically by allowing a less granular means of implementation. The enforcement of any Net Neutrality regulations is unlikely to be affected by content being encrypted. The IETF's Policy on Wiretapping can be found in [RFC2804], which does not support wiretapping in standards.

#### 2.1.7.4. Application Layer Gateways

The policy of some mobile network service providers to deploy Application Layer Gateways (ALG). Section 2.9 of [RFC2663] describes the role of ALG and their interaction with NAT and/or the application payload. ALG are deployed to provide connectivity across Network Address Translators (NAT), Firewalls, and/or Load Balancers for specific applications the mobile network providers choose to support. One example is a video application that uses the Real Time Session Protocol (RTSP) [RFC2326] primary stream as a means to identify related Real Time Protocol/Real Time Control Protocol (RTP/RTCP) [RFC3550] flows at set-up. The ALG relies on the 5-tuple flow information derived from RTSP to provision NAT or other middle boxes and provide connectivity. Implementations vary, and two examples follow:

1. Parse the content of the RTSP stream and identify the 5-tuple of the supporting streams as they are being negotiated.

2. Intercept and modify the 5-tuple information of the supporting media streams as they are being negotiated on the RTSP stream, which is more intrusive to the media streams.

#### 2.1.7.5. HTTP Header Enrichment

HTTP header enrichment (see section 3.2.1 of [RFC7230]) has been a mechanism for the mobile carrier to provide "allowed" (Non-Customer Proprietary Network Information) subscriber information to third parties or other internal systems [Enrich]. Third parties can in turn provide customized service, or use it to bill the customer or allow/block selective content. This header-enrichment method is also used within the mobile network service provider to pass information internally between sub-systems, thus keeping the internal systems loosely-coupled. With encryption, the mobile network service

provider loses the capability to include any information in the content itself.

### 2.2. Network Monitoring for Performance Management and Troubleshooting

Similar to DPI, the performance of some services can be more efficiently managed and repaired when information on user transactions is available to the service provider. It may be possible to continue such monitoring activities without clear text access to the application layers of interest, but inaccuracy will increase and efficiency of repair activities will decrease. For example, an application protocol error or failure would be opaque to network troubleshooters when transport encryption is applied, making root cause location more difficult and therefore increasing the time-to-repair. Repair time directly reduces the availability of the service, and availability is a key metric for Service Level Agreements and subscription rebates. Also, there may be more cases of user communication failures when the additional encryption processes are introduced, leading to more customer service contacts and (at the same time) less information available to network operations repair teams.

With the growing use of WebSockets [RFC6455], many forms of communications (from isochronous/real-time to bulk/elastic file transfer) will take place over HTTP port 80 or port 443, so only the messages and higher-layer data will make application differentiation possible. If the monitoring systems sees only "HTTP port 443", it cannot distinguish application streams that would benefit from priority queueing from others that would not.

## 3. Encryption in Hosting SP Environments

Hosted environments have had varied requirements in the past for encryption, with many businesses choosing to use these services primarily for data and applications that are not business or privacy sensitive. A shift prior to the revelations on surveillance/passive monitoring began where businesses were asking for hosted environments to provide higher levels of security so that additional applications and service could be hosted externally. Businesses understanding the threats of monitoring in hosted environments only increased that pressure to provide more secure access and session encryption to protect the management of hosted environments as well as for the data and applications.

### 3.1. Management Access Security

Hosted environments may have multiple levels of management access, where some may be strictly for the Hosting SP (infrastructure that may be shared among customers) and some may be accessed by a specific customer for application management. In some cases, there are multiple levels of hosting service providers, further complicating the security of management infrastructure and the associated requirements.

Hosting service provider management access is typically segregated from other traffic with a control channel and may or may not be encrypted depending upon the isolation characteristics of the management session. Customer access may be through a dedicated connection, but discussion for that connection method is out-of-scope.

#### 3.1.1. Customer Access Monitoring

Hosted applications that allow some level of customer management access may also require monitoring by the hosting service provider. The monitoring needs could include access control restrictions such as authentication, authorization, and accounting for filtering and firewall rules to ensure they are continuously met. Customer access may occur on multiple levels, including user-level and administrative access. The hosting service provider may need to monitor access either through session monitoring or log evaluation to ensure security service level agreements (SLA) for access management are met. The use of session encryption to access hosted environments limits access restrictions to the metadata described below. Monitoring and filtering may occur at an:

2-tuple IP-level with source and destination IP addresses alone, or

5-tuple IP and protocol-level with source IP address, destination IP address, protocol number, source port number, and destination port number.

Session encryption at the application level, TLS for example, currently allows access to the 5-tuple. IP-level encryption, such as IPsec in tunnel mode prevents access to the 5-tuple and may limit the ability to restrict traffic via filtering techniques. This shift may not impact all hosting service provider solutions as alternate controls may be used to authenticate sessions or access may require that clients access such services by first connecting to the organization before accessing the hosted application. Shifts in access may be required to maintain equivalent access control management. Logs may also be used for monitoring access control

restrictions are met, but would be limited to the data that could be observed due to encryption at the point of log generation. Log analysis is out of scope for this document.

### 3.1.2. Application SP Content Monitoring

The following observations apply to any IT organization that is responsible for delivering services, whether to third-parties, for example as a web based service, or to internal customers in an enterprise, e.g. a data processing system that forms a part of the enterprise's business.

Organizations responsible for the operation of a data center have many processes which access the contents of IP packets (passive methods of measurement, as defined in [RFC7799]). These processes are typically for service assurance or security purposes and form an integral and mission-critical part of data center operations.

Examples include:

- Network Performance Monitoring/Application Performance Monitoring

- Intrusion defense/prevention systems

- Malware detection

- Fraud Monitoring

- Application DDOS protection

- Cyber-attack investigation

- Proof of regulatory compliance

Many application service providers simply terminate sessions to/from the Internet at the edge of the data center in the form of SSL/TLS offload in the load balancer. Not only does this reduce the load on application servers, it simplifies the processes listed above.

However, in some situations, encryption deeper in the data center may be necessary to protect personal information or in order to meet industry regulations, e.g. those set out by the Payment Card Industry (PCI). In such situations, various methods can be used to allow trusted service assurance and security processes to access unencrypted data. These include SSL/TLS decryption in dedicated units, which then forward packets to trusted tools, or by real-time or post-capture decryption in the tools themselves.

Data center operators may also maintain packet recordings in order to be able to investigate attacks, breach of internal processes, etc. In some industries, organizations may be legally required to maintain such information for compliance purposes. Investigations of this nature require access to the unencrypted contents of the packet.

Application Service Providers may offer content-level monitoring options to detect intellectual property leakage, or other attacks. The use of session encryption will prevent Data Leakage Protection (DLP) used on the session streams from accessing content to search on keywords or phases to detect such leakage. DLP is often used to prevent the leakage of Personally Identifiable Information (PII) as well as financial account information, Personal Health Information (PHI), and Payment Card Information (PCI). If session encryption is terminated at a gateway prior to accessing these services, DLP on session data can still be performed. The decision of where to terminate encryption to hosted environments will be a risk decision made between the application service provider and customer organization according to their priorities. DLP can be performed at the server for the hosted application and on an end users system in an organization as alternate or additional monitoring points of content, however this is not frequently done in a service provider environment.

Secure overlay networks (for example, VXLAN) may be used in multi-tenancy scenarios to provide isolation assurance and thwart some active attacks. Section 7 of [RFC7348] describes some of the security issues possible when deploying VXLAN on Layer 2 networks. Rogue endpoints can join the multicast groups that carry broadcast traffic, for example. Tunneled traffic on VXLAN can be secured by using IPsec, but this adds the requirement for authentication infrastructure and may reduce packet transfer performance. Deployment of data path acceleration technologies can help to mitigate the performance issues, but they also bring more complex networking and management.

## 3.2. Hosted Applications

Organizations are increasingly using hosted applications rather than in house solutions that require maintenance of equipment and software. Examples include Enterprise Resource Planning (ERP) solutions, payroll service, time and attendance, travel and expense reporting among others. Organizations may require some level of management access to these hosted applications and will typically require session encryption or a dedicated channel for this activity.

In other cases, hosted applications may be fully managed by a hosting service provider with service level agreement expectations for

availability and performance as well as for security functions including malware detection.

### 3.2.1. Monitoring needs for Managed Applications

Performance, availability, and other aspects of a SLA are often collected through passive monitoring. For example:

- Availability: ability to establish connections with hosts to access applications, and discern the difference between network or host-related causes of unavailability.

- Performance: ability to complete transactions within a target response time, and discern the difference between network or host-related causes of excess response time.

Here, as with all passive monitoring, the accuracy of inferences are dependent on the cleartext information available, and encryption would tend to reduce the information and therefore, the accuracy of each inference. Passive measurement of some metrics will be impossible with encryption that prevents inferring packet correspondence across multiple observation points, such as for packet loss metrics.

### 3.2.2. Mail Service Providers

Mail (application) service providers vary in what services they offer. Options may include a fully hosted solution where mail is stored external to an organization's environment on mail service provider equipment or the service offering may be limited to monitor incoming mail to remove SPAM [Section 5.1], malware [Section 5.6], and phishing attacks [Section 5.3] before mail is directed to the organization's equipment. In both of these cases, content of the messages and headers is monitored to detect SPAM, malware, phishing, and other messages that may be considered an attack.

STARTTLS ought have zero effect on anti-SPAM efforts for SMTP traffic. Anti-SPAM services could easily be performed on an SMTP gateway, eliminating the need for TLS decryption services.

Many efforts are emerging to improve user-to-user encryption to protect end user's privacy. There are no clear front runners with efforts ranging from proprietary to open source ones like "Dark Mail".

### 3.3. Data Storage

Numerous service offerings exist that provide hosted storage solutions. This section describes the various offerings and details the monitoring for each type of service and how encryption may impact the operational and security monitoring performed.

Trends in data storage encryption for hosted environments include a range of options. The following list is intentionally high-level to describe the types of encryption used in coordination with data storage that may be hosted remotely, meaning the storage is physically located in an external data center requiring transport over the Internet. Options for monitoring will vary with both approaches from what may be done today.

#### 3.3.1. Host-level Encryption

For higher security and/or privacy of data and applications, options that provide end-to-end encryption of the data from the users desktop or server to the storage platform may be preferred. With this description, host level encryption includes any solution that encrypts data at the object level, not transport level. Encryption of data may be performed with libraries on the system or at the application level, which includes file encryption services via a file manager. Host-level encryption is useful when data storage is hosted, or scenarios when storage location is determined based on capacity or based on a set of parameters to automate decisions. This could mean that large data sets accessed infrequently could be sent to an off-site storage platform at an external hosting service, data accessed frequently may be stored locally, or the decision could be based on the transaction type. Host-level encryption is grouped separately for the purpose of this document as the monitoring needs as this data can be stored in multiple locations including off-site remote storage platforms. If session encryption is used, the protocol is likely to be TLS.

##### 3.3.1.1. Monitoring for Hosted Storage

The general monitoring needs of hosted storage solutions that use host-level (object) encryption is described in this subsection. Solutions might include backup services and external storage services, such as those that burst data that exceeds internal limits on occasion to external storage platforms operated by a third party.

Monitoring of data flows to hosted storage solutions is performed for security and operational purposes. The security monitoring may be to detect anomalies in the data flows that could include changes to destination, the amount of data transferred, or alterations in the

size and frequency of flows. Operational considerations include capacity and availability monitoring.

### 3.3.2. Disk Encryption, Data at Rest

There are multiple ways to achieve full disk encryption for stored data. Encryption may be performed on data to be stored while in transit close to the storage media with solutions like Controller Based Encryption (CBE) or in the drive system with Self-Encrypting Drives (SED). Session encryption is typically coupled with encryption of these data at rest (DAR) solutions to also protect data in transit. Transport encryption is likely via TLS.

#### 3.3.2.1. Monitoring Session Flows for DAR Solutions

The general monitoring needs for transport of data to storage platforms, where object level encryption is performed close to or on the storage platform are similar to those described in the section on Monitoring for Hosted Storage. The primary difference for these solutions is the possible exposure of sensitive information, which could include privacy related data, financial information, or intellectual property if session encryption via TLS is not deployed. Session encryption is typically used with these solutions, but that decision would be based on a risk assessment. There are use cases where DAR or disk-level encryption is required. Examples include preventing exposure of data if physical disks are stolen or lost.

### 3.3.3. Cross Data Center Replication Services

Storage services also include data replication which may occur between data centers and may leverage Internet connections to tunnel traffic. The traffic may use iSCSI [RFC7143] or FC/IP [RFC7146] encapsulated in IPsec. Either transport or tunnel mode may be used for IPsec depending upon the termination points of the IPsec session, if it is from the storage platform itself or from a gateway device at the edge of the data center respectively.

#### 3.3.3.1. Monitoring Of IPSec for Data Replication Services

The general monitoring needs for data replication are described in this subsection.

Monitoring of data flows between data centers may be performed for security and operational purposes and would typically concentrate more on the operational needs since these flows are essentially virtual private networks (VPN) between data centers. Operational considerations include capacity and availability monitoring. The security monitoring may be to detect anomalies in the data flows,

similar to what was described in the "Monitoring for Hosted Storage Section".

## 4. Encryption for Enterprises

Encryption of network traffic within the private enterprise is a growing trend, particularly in industries with audit and regulatory requirements. Some enterprise internal networks are almost completely TLS and/or IPsec encrypted.

For each type of monitoring, different techniques and access to parts of the data stream are part of current practice. As we transition to an increased use of encryption, alternate methods of monitoring for operational purposes may be necessary to reduce the need to break encryption and thus privacy of users (other policies may apply in some enterprise settings).

### 4.1. Monitoring Needs of the Enterprise

Large corporate enterprises are the owners of the platforms, data, and network infrastructure that provide critical business services to their user communities. As such, these enterprises are responsible for all aspects of the performance, availability, security, and quality of experience for all user sessions. These responsibilities break down into three basic areas:

1. Security Monitoring and Control

2. Application Performance Monitoring and Reporting

3. Network Diagnostics and Troubleshooting

In each of the above areas, technical support teams utilize collection, monitoring, and diagnostic systems. Some organizations currently use attack methods such as replicated TLS server RSA private keys to decrypt passively monitored copies of encrypted TLS packet streams.

For an enterprise to avoid costly application down time and deliver expected levels of performance, protection, and availability, some forms of traffic analysis sometimes including examination of packet payloads are currently used.

#### 4.1.1. Security Monitoring in the Enterprise

Enterprise users are subject to the policies of their organization and the jurisdictions in which the enterprise operates. As such, proxies may be in use to:

1. intercept outbound session traffic to monitor for intellectual property leakage (by users or more likely these days through malware and trojans),

2. detect viruses/malware entering the network via email or web traffic,

3. detect malware/Trojans in action, possibly connecting to remote hosts,

4. detect attacks (Cross site scripting and other common web related attacks),

5. track misuse and abuse by employees,

6. restrict the types of protocols permitted to/from the entire corporate environment,

7. detect and defend against Internet DDoS attacks, including both volumetric and layer 7 attacks.

A significant portion of malware hides its activity within TLS or other encrypted protocols. This includes lateral movement, Command and Control, and Data Exfiltration. Detecting these functions are important to effective monitoring and mitigation of malicious traffic, not limited to malware.

### 4.1.2. Application Performance Monitoring in the Enterprise

There are two main goals of monitoring:

1. Assess traffic volume on a per-application basis, for billing, capacity planning, optimization of geographical location for servers or proxies, and other needs.

2. Assess performance in terms of application response time and user perceived response time.

Network-based Application Performance Monitoring tracks application response time by user and by URL, which is the information that the application owners and the lines of business need. Content Delivery Networks (CDNs) add complexity in determining the ultimate endpoint destination. By their very nature, such information is obscured by CDNs and encrypted protocols -- adding a new challenge for troubleshooting network and application problems. URL identification allows the application support team to do granular, code level troubleshooting at multiple tiers of an application.

New methodologies to monitor user perceived response time and to separate network from server time are evolving. For example, the IPv6 Destination Option Header (DOH) implementation of Performance and Diagnostic Metrics (PDM) will provide this [I-D.ietf-ippm-6man-pdm-option]. Using PDM with IPSec Encapsulating Security Payload (ESP) Transport Mode requires placement of the PDM DOH within the ESP encrypted payload to avoid leaking timing and sequence number information that could be useful to an attacker. Use of PDM DOH also may introduce some security weaknesses, including a timing attack, as described in Section 7 of [I-D.ietf-ippm-6man-pdm-option]. For these and other reasons, [I-D.ietf-ippm-6man-pdm-option] requires that the PDM DOH option be explicitly turned on by administrative action in each host where this measurement feature will be used.

### 4.1.3. Enterprise Network Diagnostics and Troubleshooting

One primary key to network troubleshooting is the ability to follow a transaction through the various tiers of an application in order to isolate the fault domain. A variety of factors relating to the structure of the modern data center and the modern multi-tiered application have made it difficult to follow a transaction in network traces without the ability to examine some of the packet payload. Alternate methods, such as log analysis need improvement to fill this gap.

#### 4.1.3.1. NAT

Content Delivery Networks (CDNs) and NATs obscure the ultimate endpoint designation. Troubleshooting a problem for a specific end user requires finding information such as the IP address and other identifying information so that their problem can be resolved in a timely manner.

NAT is also frequently used by lower layers of the data center infrastructure. Firewalls, Load Balancers, Web Servers, App Servers, and Middleware servers all regularly NAT the source IP of packets. Combine this with the fact that users are often allocated randomly by load balancers to all these devices, the network troubleshooter is often left with very few options in today's environment due to poor logging implementations in applications. As such, network troubleshooting is used to trace packets at a particular layer, decrypt them, and look at the payload to find a user session.

This kind of bulk packet capture and bulk decryption is frequently used when troubleshooting a large and complex application. Endpoints typically don't have the capacity to handle this level of network packet capture, so out-of-band networks of robust packet brokers and

network sniffers that use techniques such as copies of TLS RSA private keys accomplish this task today.

#### 4.1.3.2. TCP Pipelining/Session Multiplexing

TCP Pipelining/Session Multiplexing used mainly by middle boxes today allow for multiple end user sessions to share the same TCP connection. Today's network troubleshooter often relies upon session decryption to tell which packet belongs to which end user as the logs are currently inadequate for the analysis needed.

With the advent of HTTP2, session multiplexing will be used ubiquitously, both on the Internet and in the private data center.

#### 4.1.3.3. HTTP Service Calls

When an application server makes an HTTP service call to back end services on behalf of a user session, it uses a completely different URL and a completely different TCP connection. Troubleshooting via network trace involves matching up the user request with the HTTP service call. Some organizations do this today by decrypting the TLS packet and inspecting the payload. Logging has not been adequate for their purposes.

#### 4.1.3.4. Application Layer Data

Many applications use text formats such as XML to transport data or application level information. When transaction failures occur and the logs are inadequate to determine the cause, network and application teams work together, each having a different view of the transaction failure. Using this troubleshooting method, the network packet is correlated with the actual problem experienced by an application to find a root cause. The inability to access the payload prevents this method of troubleshooting.

### 4.2. Techniques for Monitoring Internet Session Traffic

Corporate networks commonly monitor outbound session traffic to detect or prevent attacks as well as to guarantee service level expectations. In some cases, alternate options are available when encryption is in use, but techniques like that of data leakage prevention tools at a proxy would not be possible if encrypted traffic can not be intercepted, encouraging alternate options such as performing these functions at the edge.

Data leakage detection prevention (DLP) tools intercept traffic at the Internet gateway or proxy services with the ability to man-in-the-middle (MiTM) encrypted session traffic (HTTP/TLS). These tools

may use key words important to the enterprise including business sensitive information such as trade secrets, financial data, personally identifiable information (PII), or personal health information (PHI). Various techniques are used to intercept HTTP/TLS sessions for DLP and other purposes, and are described in "Summarizing Known Attacks on TLS and DTLS" [RFC7457]. Note: many corporate policies allow access to personal financial and other sites for users without interception.

Monitoring traffic patterns for anomalous behavior such as increased flows of traffic that could be bursty at odd times or flows to unusual destinations (small or large amounts of traffic). This traffic may or may not be encrypted and various methods of encryption or just obfuscation may be used.

Restrictions on traffic to approved sites: Web proxies are sometimes used to filter traffic, allowing only access to well-known sites found to be legitimate and free of malware on last check by a proxy service company. This type of restriction is usually not noticeable in a corporate setting, but may be to those in research who are unable to access colleague's individual sites or new web sites that have not yet been screened. In situations where new sites are required for access, they can typically be added after notification by the user or proxy log alerts and review. Home mail account access may be blocked in corporate settings to prevent another vector for malware to enter as well as for intellectual property to leak out of the network. This method remains functional with increased use of encryption and may be more effective at preventing malware from entering the network. Web proxy solutions monitor and potentially restrict access based on the destination URL or the DNS name. A complete URL may be used in cases where access restrictions vary for content on a particular site or for the sites hosted on a particular server.

Desktop DLP tools are used in some corporate environments as well. Since these tools reside on the desktop, they can intercept traffic before it is encrypted and may provide a continued method of monitoring intellectual property leakage from the desktop to the Internet or attached devices.

DLP tools can also be deployed by Network Service providers, as they have the vantage point of monitoring all traffic paired with destinations off the enterprise network. This makes an effective solution for enterprises that allow "bring-your-own" devices when the traffic is not encrypted and devices that do not fit the desktop category, but are used on corporate networks nonetheless.

Enterprises may wish to reduce the traffic on their Internet access facilities by monitoring requests for within-policy content and caching it. In this case, repeated requests for Internet content spawned by URLs in e-mail trade newsletters or other sources can be served within the enterprise network. Gradual deployment of end to end encryption would tend to reduce the cacheable content over time, owing to concealment of critical headers and payloads. Many forms of enterprise performance management and optimization based on monitoring (DPI) would suffer the same fate.

## 5. Security Monitoring for Specific Attack Types

Effective incident response today requires collaboration at Internet scale. This section will only focus on efforts of collaboration at Internet scale that are dedicated to specific attack types. They may require new monitoring and detection techniques in an increasingly encrypted Internet. As mentioned previously, some service providers have been interfering with STARTTLS to prevent session encryption to be able to perform functions they are used to (injecting ads, monitoring, etc.). By detailing the current monitoring methods used for attack detection and response, this information can be used to devise new monitoring methods that will be effective in the changed Internet via collaboration and innovation.

### 5.1. Mail Abuse and SPAM

The largest operational effort to prevent mail abuse is through the Messaging, Malware, Mobile Anti-Abuse Working Group (M3AAWG)[M3AAWG]. Mail abuse is combated directly with mail administrators who can shut down or stop continued mail abuse originating from large scale providers that participate in using the Abuse Reporting Format (ARF) agents standardized in the IETF [RFC5965], [RFC6430], [RFC6590], [RFC6591], [RFC6650], [RFC6651], and [RFC6652]. The ARF agent directly reports abuse messages to the appropriate service provider who can take action to stop or mitigate the abuse. Since this technique uses the actual message, the use of SMTP over TLS between mail gateways will not effect its usefulness. As mentioned previously, SMTP over TLS only protects data while in transit and the messages may be exposed on mail servers or mail gateways if a user-to-user encryption method is not used. Current user-to-user message encryption methods on email (S/MIME and PGP) do not encrypt the email header information used by ARF and the service provider operators in their abuse mitigation efforts.

### 5.2. Denial of Service

Response to Denial of Service (DoS) attacks are typically coordinated by the SP community with a few key vendors who have tools to assist in the mitigation efforts. Traffic patterns are determined from each DoS attack to stop or rate limit the traffic flows with patterns unique to that DoS attack.

Data types used in monitoring traffic for DDoS are described in the DDoS Open Threat Signaling (DOTS) working group documents in development.

Data types used in DDoS attacks have been detailed in the IODEF Guidance draft [I-D.ietf-mile-iodef-guidance], Appendix A.2, with the help of several members of the service provider community. The examples provided are intended to help identify the useful data in detecting and mitigating these attacks independent of the transport and protocol descriptions in the drafts.

### 5.3. Phishing

Investigations and response to phishing attacks follow well-known patterns, requiring access to specific fields in email headers as well as content from the body of the message. When reporting phishing attacks, the recipient has access to each field as well as the body to make content reporting possible, even when end-to-end encryption is used. The email header information is useful to identify the mail servers and accounts used to generate or relay the attack messages in order to take the appropriate actions. The content of the message often contains an embedded attack that may be in an infected file or may be a link that results in the download of malware to the users system.

Administrators often find it helpful to use header information to track down similar message in their mail queue or users inboxes to prevent further infection. Combinations of To:, From:, Subject:, Received: from header information might be used for this purpose. Administrators may also search for document attachments of the same name, size, or containing a file with a matching hash to a known phishing attack. Administrators might also add URLs contained in messages to block lists locally or this may also be done by browser vendors through larger scales efforts like that of the Anti-Phishing Working Group (APWG).

A full list of the fields used in phishing attack incident response can be found in RFC5901. Future plans to increase privacy protections may limit some of these capabilities if some email header fields are encrypted, such as To:, From:, and Subject: header fields.

This does not mean that those fields should not be encrypted, only that we should be aware of how they are currently used. Alternate options to detect and prevent phishing attacks may be needed. More recent examples of data exchanged in spear phishing attacks has been detailed in the IODEF Guidance draft [I-D.ietf-mile-iodef-guidance], Appendix A.3.

## 5.4. Botnets

Botnet detection and mitigation is complex and may involve hundreds or thousands of hosts with numerous Command and Control (C&C) servers. The techniques and data used to monitor and detect each may vary. Connections to C&C servers are typically encrypted, therefore a move to an increasingly encrypted Internet may not affect the detection and sharing methods used.

## 5.5. Malware

Malware monitoring and detection techniques vary. As mentioned in the enterprise section, malware monitoring may occur at gateways to the organization analyzing email and web traffic. These services can also be provided by service providers, changing the scale and location of this type of monitoring. Additionally, incident responders may identify attributes unique to types of malware to help track down instances by their communication patterns on the Internet or by alterations to hosts and servers.

Data types used in malware investigations have been summarized in an example of the IODEF Guidance draft [I-D.ietf-mile-iodef-guidance], Appendix A.1.

## 5.6. Spoofed Source IP Address Protection

The IETF has reacted to spoofed source IP address-based attacks, recommending the use of network ingress filtering [RFC2827] and the unicast Reverse Path Forwarding (uRPF) mechanism [RFC2504]. But uRPF suffers from limitations regarding its granularity: a malicious node can still use a spoofed IP address included inside the prefix assigned to its link. The Source Address Validation Improvements (SAVI) mechanisms try to solve this issue. Basically, a SAVI mechanism is based on the monitoring of a specific address assignment/management protocol (e.g., SLAAC [RFC4682], SEND [RFC3791], DHCPv4/v6 [RFC2131][RFC3315]) and, according to this monitoring, set-up a filtering policy allowing only the IP flows with a correct source IP address (i.e., any packet with a source IP address, from a node not owning it, is dropped). The encryption of parts of the address assignment/management protocols, critical for SAVI mechanisms, can result in a dysfunction of the SAVI mechanisms.

### 5.7. Further work

Although incident response work will continue, new methods to prevent system compromise through security automation and continuous monitoring [SACM] may provide alternate approaches where system security is maintained as a preventative measure.

## 6. Application-based Flow Information Visible to a Network

This section describes specific techniques used in monitoring applications that may apply to various network types.

### 6.1. TLS Server Name Indication

When initiating the TLS handshake, the Client may provide an extension field (server_name) which indicates the server to which it is attempting a secure connection. TLS SNI was standardized in 2003 to enable servers to present the "correct TLS certificate" to clients in a deployment of multiple virtual servers hosted by the same server infrastructure and IP-address. Although this is an optional extension, it is today supported by all modern browsers, web servers and developer libraries. It should be noted that HTTP/2 introduces the Alt-SVC method for upgrading the connection from HTTP/1 to either unencrypted or encrypted HTTP/2. If the initial HTTP/1 request is unencrypted, the destination alternate service name can be identified before the communication is potentially upgraded to encrypted HTTP/2 transport. HTTP/2 implementations MUST support the Server Name Indication (SNI) extension.

This information is only visible if the client is populating the Server Name Indication extension. This need not be done, but may be done as per TLS standard. Therefore, even if existing network filters look out for seeing a Server Name Indication extension, they may not find one. The per-domain nature of SNI may not reveal the specific service or media type being accessed, especially where the domain is of a provider offering a range of email, video, Web pages etc. For example, certain blog or social network feeds may be deemed 'adult content', but the Server Name Indication will only indicate the server domain rather than a URL path.

### 6.2. Application Layer Protocol Negotiation (ALPN)

ALPN is a TLS extension which may be used to indicate the application protocol within the TLS session. This is likely to be of more value to the network where it indicates a protocol dedicated to a particular traffic type (such as video streaming) rather than a multi-use protocol. ALPN is used as part of HTTP/2 'h2', but will

not indicate the traffic types which may make up streams within an HTTP/2 multiplex.

### 6.3. Content Length, BitRate and Pacing

The content length of encrypted traffic is effectively the same as the cleartext. Although block ciphers utilise padding this makes a negligible difference. Bitrate and pacing are generally application specific, and do not change much when the content is encrypted. Multiplexed formats (such as HTTP/2 and QUIC) may however incorporate several application streams over one connection, which makes the bitrate/pacing no longer application-specific.

## 7. Response to Increased Encryption and Looking Forward

In the best case scenario, engineers and other innovators would work to solve the problems at hand in new ways rather than prevent the use of encryption. It will take time to devise alternate approaches to achieve similar goals.

There has already been documented cases of service providers preventing STARTTLS [NoEncrypt] to prevent session encryption negotiation on some session to inject a super cookie.

National surveillance programs have a clear need for monitoring terrorism [JNSLP] as do Internet security practitioners monitoring for criminal activities. Governments vary on their balance between their need for monitoring versus the protection of user privacy, data, and assets. Those that favor unencrypted access to data ignore the real need to protect users identity, financial transactions and intellectual property, which requires security and encryption to prevent crime. A clear understanding of technology, encryption, and monitoring needs will aid in the development of solutions to appropriately balance the need of privacy. As this understanding increases, hopefully the discussions will improve and this draft is meant to help further the discussion.

Terrorists and criminals have been using encryption for many years. The current push to increase encryption is aimed at increasing users privacy. There is already protection in place for purchases, financial transactions, systems management infrastructure, and intellectual property although this too can be improved. The Opportunistic Security (OS) [RFC7435] efforts aim to increase the costs of monitoring through the use of encryption that can be subject to active attacks, but make passive monitoring broadly cost prohibitive. This is meant to restrict monitoring to sessions where there is reason to have suspicion.

Open questions: As the use of encryption increases, does passive monitoring become limited to metadata analysis? What metadata should be left in protocols as they evolve to also protect users privacy? Can we make changes to protocols and message exchanges to alter the current monitoring needs at least for operations and security practitioners?

Options are on the technology horizon that could help to end the struggle between the need to monitor by operators, security teams, and nations and those seeking to protect users privacy if they come to fruition. The solutions are very interesting, but are at least several years out and include homomorphic encrypt, functional encryption, and filterable decryption [homomorphic]. This technology will allow for searching and pattern matching on encrypted data, but is still several years out.

## 8. Security Considerations

There are no additional security considerations as this is a summary and does not include a new protocol or functionality.

## 9. IANA Considerations

This memo makes no requests of IANA.

## 10. Acknowledgements

Thanks to our reviewers, Natasha Rooney, Kevin Smith, Ashutosh Dutta, Brandon Williams, Jean-Michel Combes, Nalini Elkins, Paul Barrett, and Stephen Farrell for their editorial and content suggestions. Surya K. Kovvali provided material for the Appendix.

## 11. Appendix: Impact on Mobility Network Optimizations and New Services

This Appendix considers the effects of transport level encryption on existing forms of mobile network optimization techniques, as well as potential new services. The material in this Appendix assumes familiarity with mobile network concepts, specifications, and architectures. Readers who need additional background should start with the 3GPP's web pages on various topics of interest[Web3GPP], especially the article on LTE. 3GPP provides a mapping between their expanding technologies and the different series of technical specifications [Map3GPP]. 3GPP also has a canonical specification of their vocabulary, definitions, and acronyms [Vocab], as does the RFC Editor for abbreviations [RFCEdit].

### 11.1. Effect of Encypted ACKs

The stream of TCP ACKs that flow from a receiver of a byte stream using TCP for reliability, flow-control, and NAT/firewall transversal is called an ACK stream. The ACKs contain segment numbers that confirm successful transmission and their RTT, or indicate packet loss (duplicate ACKs). If this view of progress of stream transfer is lost, then the mobile network has greatly reduced ability to monitor transport layer performance. When the ACK stream is encrypted, it prevents the following mobile network features from operating:

a. Measurement of Network Segment (Sector, eNodeB (eNB) etc.) characterization KPIs (Retransmissions, packet drops, Sector Utilization Level etc.), estimation of User/Service KQIs at network edges for circuit emulation (CEM), and mitigation methods. The active services per user and per sector are not visible to a server that only services Internet Access Point Names (APN), and thus could not perform mitigation functions based on network segment view.

b. Retransmissions by trusted proxies at network edges that improve live transmission over long delay, capacity-varying networks.

c. Content replication near the network edge (for example live video, DRM protected content) to maximize QOE. Replicating every stream through the transit network increases backhaul cost for live TV.

d. Ability to deploy trusted proxies that reduce control round-trip time (RTT) between the TCP transmitter and receiver. The RTT determines how quickly a user's attempt to cancel a video is recognized (how quickly the traffic is stopped, thus keeping un-wanted video packets from entering the radio scheduler queue).

e. Trusted proxy with low RTT determines the responsiveness of TCP flow control, and enables faster adaptation in a delay & capacity varying network due to user mobility. Low RTT permits use of a smaller send window, which makes the flow control loop more responsive to changing mobile network conditions.

f. Opportunistic RAN-Aware pacing, acceleration, and deferral of high volume content such as video or software updates.

### 11.2. Effect of Encrypted Transport Headers

When the Transport Header is encrypted, it prevents the following mobile network features from operating:

a. Application-type-aware network edge (middlebox) that could control pacing, limit simultaneous HD videos, prioritize active videos against new videos, etc.

b. For the Access Network Discovery and Selection Function (3GPP-ANDSF), it Impedes content-aware network selection for steering users or specific flows to appropriate Networks.

c. For Self Organizing Networks (3GPP SON) - intelligent SON workflows such as content-ware MLB (Mobility Load Balancing)

d. For User Plane Congestion Management (3GPP UPCON) - ability to understand content and manage network during congestion. Mitigating techniques such as deferred download, off-peak acceleration, and outbound roamers.

e. Reduces the benefits IP/DSCP-based transit network delivery optimizations; since the multiple applications are multiplexed within the same 5-tuple transport connection, the DSCP markings would not correspond to an application flow.

f. Advance notification for dense data usages - If the application types are visible, transit network element could warn (ahead of usage) that the requested service consumes user plan limits, and transmission could be terminated. Without such visibility the network might have to continue the operation and stop the operation after the limit, because partially loaded content wastes resources and may not be usable by the client thus increasing customer complaints. Content publisher will not know user-service plans, and Network Edge would not know data transfer lengths before large object is requested.

### 11.3. Effect of Encryption on New Services

This section describes some new mobile services and how they might be affected with transport encryption:

1. Flow-based charging allowing zero-rated and monetized traffic; for example operators may charge nothing, or charge based on domain/URLs.

2. Content/Application based Prioritization of Over-the-Top (OTT) services - each application-type or service has different

delay/loss/throughput expectations, and each type of stream will be unknown to an edge device if encrypted; this impedes dynamic-QoS adaptation.

3. Rich Communication Services (3GPP-RCS) using different Quality Class Indicators (QCIs in LTE) - Operators offer different QoS classes for value-added services. The QCI type is visible in RAN control plane and invisible in user plane, thus the QCI cannot be set properly when the application -type is unknown.

4. Enhanced Multimedia Broadcast/Multicast Services (3GPP eMBMS) - trusted edge proxies facilitate delivering same stream to different users, using either unicast or multicast depending on channel conditions to the user.

5. Transport level protection is unnecessary for already protected content (such as content with Digital Rights Management, DRM). It prevents multi-user replication, and tandem encryption stages increase required processing cycles.

### 11.4. Effect of Encryption on Mobile Network Evolution

The transport header encryption prevents trusted transit proxies. It may be that the benefits of such proxies could be achieved by end to end client & server optimizations and distribution using CDNs, plus the ability to continue connections across different access technologies (across dynamic user IP addresses). The following aspects need to be considered in this approach:

1. In a wireless mobile network, the delay and channel capacity per user and sector varies due to coverage, contention, user mobility, and scheduling balances fairness, capacity and service QoE. If most users are at the cell edge, the controller cannot use more complex QAM, thus reducing total cell capacity; similarly if a UMTS edge is serving some number of CS-Voice Calls, the remaining capacity for packet services is reduced.

2. Inbound Roamers: Mobile wireless networks service in-bound roamers (Users of Operator A in a foreign operator Network B) by backhauling their traffic though Operator B's network to Operator A's Network and then serving through the P-Gateway (PGW), General GPRS Support Node (GGSN), Content Distribution Network (CDN) etc., of Operator A (User's Home Operator). Increasing window sizes to compensate for the path RTT will have the limitations outlined earlier for TCP.

3. Outbound Roamers: Similar to inbound roamers, users accessing different Core/Content network, for example domains not serviced

via local CDNs are carried through operator network via different APN or GW to remote networks which increases path RTT & control loop.

4. Issues in deploying CDNs in RAN: Decreasing Client-Server control loop requires deploying CDNs/Cloud functions that terminate encryption closer to the edge. In Cellular RAN, the user IP traffic is encapsulated into GPSR Tunneling Protocol-User Plane (GTP-U in UMTS and LTE) tunnels to handle user mobility; the tunnels terminate in APN/GGSN/PGW that are in central locations. One user's traffic may flow through one or more APN's (for example Internet APN, Roaming APN for Operator X, Video-Service APN, OnDeckAPN etc.). The scope of operator private IP addresses may be limited to specific APN. Since CDNs generally operate on user IP flows, deploying them would require enhancing them with tunnel translation, etc., tunnel management functions.

5. While CDNs that de-encrypt flows or split-connection proxy (similar to split-tcp) could be deployed closer to the edges to reduce control loop RTT, with transport header encryption, such CDNs perform optimization functions only for partner client flows; thus content from Small-Medium Businesses (SMBs) would not get such CDN benefits.

6. Mobile Edge Computing (MEC) initiative to push latency sensitive functions to the edge of the network; for example a trusted proxy could facilitate services between two devices in RAN without requiring content flow through the WebServer.

## 12. Informative References

[CAIDA] "CAIDA [http://www.caida.org/data/overview/]".

[CALEA] Pub. L. No. 103-414, 108 Stat. 4279, codified at 47 USC 1001-1010, "Communications Assistance for Law Enforcement Act (CALEA)".

[EFF] "Electronic Frontier Foundation https://www.eff.org/".

[EFF2014] "EFF Report on STARTTLS Downgrade Attacks https://www.eff.org/deeplinks/2014/11/starttls-downgrade-attacks".

[Enrich] Narseo Vallina-Rodriguez, et al., , "Header Enrichment or ISP Enrichment? Emerging Privacy Threats in Mobile Networks, Hot Middlebox'15, August 17-21 2015, London, United Kingdom", 2015.

[ETSI101331]
ETSI TS 101 331 V1.1.1 (2001-08), "Telecommunications security; Lawful Interception (LI); Requirements of Law Enforcement Agencies", August 2001.

[homomorphic]
Volume 20, 2013, Pages 502-509, Complex Adaptive Systems, "Homomorphic Encryption http://www.sciencedirect.com/science/article/pii/S1877050913011101".

[I-D.ietf-ippm-6man-pdm-option]
Elkins, N., Hamilton, R., and m. mackermann@bcbsm.com, "IPv6 Performance and Diagnostic Metrics (PDM) Destination Option", draft-ietf-ippm-6man-pdm-option-08 (work in progress), February 2017.

[I-D.ietf-mile-iodef-guidance]
Kampanakis, P. and M. Suzuki, "IODEF Usage Guidance", draft-ietf-mile-iodef-guidance-07 (work in progress), November 2016.

[JNSLP] Surveillance, Vol. 8 No. 3, "10 Standards for Oversight and Transparency of National Intelligence Services http://jnslp.com/".

[M3AAWG] "Messaging, Malware, Mobile Anti-Abuse Working Group (M3AAWG) https://www.maawg.org/".

[Map3GPP] http://www.3gpp.org/technologies, "Mapping between technologies and specifications".

[NoEncrypt]
"ISPs Removing their Customers EMail Encryption https://www.eff.org/deeplinks/2014/11/starttls-downgrade-attacks/".

[RFC2131] Droms, R., "Dynamic Host Configuration Protocol", RFC 2131, DOI 10.17487/RFC2131, March 1997, <http://www.rfc-editor.org/info/rfc2131>.

[RFC2326] Schulzrinne, H., Rao, A., and R. Lanphier, "Real Time Streaming Protocol (RTSP)", RFC 2326, DOI 10.17487/RFC2326, April 1998, <http://www.rfc-editor.org/info/rfc2326>.

[RFC2504] Guttman, E., Leong, L., and G. Malkin, "Users' Security Handbook", FYI 34, RFC 2504, DOI 10.17487/RFC2504, February 1999, <http://www.rfc-editor.org/info/rfc2504>.

[RFC2663] Srisuresh, P. and M. Holdrege, "IP Network Address Translator (NAT) Terminology and Considerations", RFC 2663, DOI 10.17487/RFC2663, August 1999, <http://www.rfc-editor.org/info/rfc2663>.

[RFC2804] IAB and IESG, "IETF Policy on Wiretapping", RFC 2804, DOI 10.17487/RFC2804, May 2000, <http://www.rfc-editor.org/info/rfc2804>.

[RFC2827] Ferguson, P. and D. Senie, "Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing", BCP 38, RFC 2827, DOI 10.17487/RFC2827, May 2000, <http://www.rfc-editor.org/info/rfc2827>.

[RFC3315] Droms, R., Ed., Bound, J., Volz, B., Lemon, T., Perkins, C., and M. Carney, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)", RFC 3315, DOI 10.17487/RFC3315, July 2003, <http://www.rfc-editor.org/info/rfc3315>.

[RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, DOI 10.17487/RFC3550, July 2003, <http://www.rfc-editor.org/info/rfc3550>.

[RFC3791] Olvera, C., Nesser, P., and , "Survey of IPv4 Addresses in Currently Deployed IETF Routing Area Standards Track and Experimental Documents", RFC 3791, DOI 10.17487/RFC3791, June 2004, <http://www.rfc-editor.org/info/rfc3791>.

[RFC4682] Nechamkin, E. and J-F. Mule, "Multimedia Terminal Adapter (MTA) Management Information Base for PacketCable- and IPCablecom-Compliant Devices", RFC 4682, DOI 10.17487/RFC4682, December 2006, <http://www.rfc-editor.org/info/rfc4682>.

[RFC5965] Shafranovich, Y., Levine, J., and M. Kucherawy, "An Extensible Format for Email Feedback Reports", RFC 5965, DOI 10.17487/RFC5965, August 2010, <http://www.rfc-editor.org/info/rfc5965>.

[RFC6430] Li, K. and B. Leiba, "Email Feedback Report Type Value: not-spam", RFC 6430, DOI 10.17487/RFC6430, November 2011, <http://www.rfc-editor.org/info/rfc6430>.

[RFC6455] Fette, I. and A. Melnikov, "The WebSocket Protocol", RFC 6455, DOI 10.17487/RFC6455, December 2011, <http://www.rfc-editor.org/info/rfc6455>.

[RFC6590] Falk, J., Ed. and M. Kucherawy, Ed., "Redaction of Potentially Sensitive Data from Mail Abuse Reports", RFC 6590, DOI 10.17487/RFC6590, April 2012, <http://www.rfc-editor.org/info/rfc6590>.

[RFC6591] Fontana, H., "Authentication Failure Reporting Using the Abuse Reporting Format", RFC 6591, DOI 10.17487/RFC6591, April 2012, <http://www.rfc-editor.org/info/rfc6591>.

[RFC6650] Falk, J. and M. Kucherawy, Ed., "Creation and Use of Email Feedback Reports: An Applicability Statement for the Abuse Reporting Format (ARF)", RFC 6650, DOI 10.17487/RFC6650, June 2012, <http://www.rfc-editor.org/info/rfc6650>.

[RFC6651] Kucherawy, M., "Extensions to DomainKeys Identified Mail (DKIM) for Failure Reporting", RFC 6651, DOI 10.17487/RFC6651, June 2012, <http://www.rfc-editor.org/info/rfc6651>.

[RFC6652] Kitterman, S., "Sender Policy Framework (SPF) Authentication Failure Reporting Using the Abuse Reporting Format", RFC 6652, DOI 10.17487/RFC6652, June 2012, <http://www.rfc-editor.org/info/rfc6652>.

[RFC7143] Chadalapaka, M., Satran, J., Meth, K., and D. Black, "Internet Small Computer System Interface (iSCSI) Protocol (Consolidated)", RFC 7143, DOI 10.17487/RFC7143, April 2014, <http://www.rfc-editor.org/info/rfc7143>.

[RFC7146] Black, D. and P. Koning, "Securing Block Storage Protocols over IP: RFC 3723 Requirements Update for IPsec v3", RFC 7146, DOI 10.17487/RFC7146, April 2014, <http://www.rfc-editor.org/info/rfc7146>.

[RFC7230] Fielding, R., Ed. and J. Reschke, Ed., "Hypertext Transfer Protocol (HTTP/1.1): Message Syntax and Routing", RFC 7230, DOI 10.17487/RFC7230, June 2014, <http://www.rfc-editor.org/info/rfc7230>.

[RFC7234] Fielding, R., Ed., Nottingham, M., Ed., and J. Reschke, Ed., "Hypertext Transfer Protocol (HTTP/1.1): Caching", RFC 7234, DOI 10.17487/RFC7234, June 2014, <http://www.rfc-editor.org/info/rfc7234>.

[RFC7258] Farrell, S. and H. Tschofenig, "Pervasive Monitoring Is an Attack", BCP 188, RFC 7258, DOI 10.17487/RFC7258, May 2014, <http://www.rfc-editor.org/info/rfc7258>.

[RFC7348] Mahalingam, M., Dutt, D., Duda, K., Agarwal, P., Kreeger, L., Sridhar, T., Bursell, M., and C. Wright, "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", RFC 7348, DOI 10.17487/RFC7348, August 2014, <http://www.rfc-editor.org/info/rfc7348>.

[RFC7435] Dukhovni, V., "Opportunistic Security: Some Protection Most of the Time", RFC 7435, DOI 10.17487/RFC7435, December 2014, <http://www.rfc-editor.org/info/rfc7435>.

[RFC7457] Sheffer, Y., Holz, R., and P. Saint-Andre, "Summarizing Known Attacks on Transport Layer Security (TLS) and Datagram TLS (DTLS)", RFC 7457, DOI 10.17487/RFC7457, February 2015, <http://www.rfc-editor.org/info/rfc7457>.

[RFC7525] Sheffer, Y., Holz, R., and P. Saint-Andre, "Recommendations for Secure Use of Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS)", BCP 195, RFC 7525, DOI 10.17487/RFC7525, May 2015, <http://www.rfc-editor.org/info/rfc7525>.

[RFC7624] Barnes, R., Schneier, B., Jennings, C., Hardie, T., Trammell, B., Huitema, C., and D. Borkmann, "Confidentiality in the Face of Pervasive Surveillance: A Threat Model and Problem Statement", RFC 7624, DOI 10.17487/RFC7624, August 2015, <http://www.rfc-editor.org/info/rfc7624>.

[RFC7799] Morton, A., "Active and Passive Metrics and Methods (with Hybrid Types In-Between)", RFC 7799, DOI 10.17487/RFC7799, May 2016, <http://www.rfc-editor.org/info/rfc7799>.

[RFC7858] Hu, Z., Zhu, L., Heidemann, J., Mankin, A., Wessels, D., and P. Hoffman, "Specification for DNS over Transport Layer Security (TLS)", RFC 7858, DOI 10.17487/RFC7858, May 2016, <http://www.rfc-editor.org/info/rfc7858>.

[RFCEdit] https://www.rfc-editor.org/materials/abbrev.expansion.txt, "RFC Editor Abbreviation List".

[Vocab] https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=558, "3GPP TR 21.905 V13.1.0 (2016-06) Vocabulary for 3GPP Specifications".

[Web3GPP] http://www.3gpp.org/technologies/95-keywords-acronyms, "3GPP Web pages on specific topics of interest".

[WebCache] Xing Xu, et al., , "Investigating Transparent Web Proxies in Cellular Networks, Passive and Active Measurement Conference (PAM)", 2015.

## Authors' Addresses

Kathleen Moriarty
Dell EMC
176 South St
Hopkinton, MA
USA

Phone: +1
Email: Kathleen.Moriarty@dell.com

Al Morton
AT&T Labs
200 Laurel Avenue South
Middletown,, NJ 07748
USA

Phone: +1 732 420 1571
Fax: +1 732 368 1192
Email: acmorton@att.com
URI: http://home.comcast.net/~acmacm/