Cross-Border Data Transfers in Education:

Investigating the Transmission of Children's Internet Data from Danish Schools to U.S. Data Centers via Google Chromebooks

Introduction

The rapid growth of digital technology and its integration into educational environments has introduced both benefits and risks, particularly concerning data privacy. In this investigation, we focus on data transfers associated with Chromebooks distributed to schoolchildren in Denmark and the potential transmission of sensitive personal data, including internet habits, across international borders. The core of this inquiry is to examine whether Google, via its Chromebook devices, transfers and stores children's data in data centers located in the United States. Such practices, if confirmed, could have significant implications under both European and international data privacy laws, including the General Data Protection Regulation (GDPR).

The specific focus of this report is to analyze network traffic captured during the use of these Chromebooks, using tools such as Wireshark and Tshark, to determine the flow of data packets and the geographical locations of the servers involved. The traffic capture will serve as primary evidence to establish whether personal data related to minors, including browsing habits, is being transmitted from Europe to data centers outside the European Union, particularly in the United States.

Given the sensitive nature of children's data and the strict regulatory frameworks governing its use, this report aims to present its findings in a manner suitable for legal scrutiny. Each section will provide clear, evidence-based conclusions derived from the network traffic analysis. Where applicable, the data flows identified will be linked to relevant legal standards, ensuring that this document serves as a reliable foundation for legal proceedings or regulatory action.

This investigation will first detail the technical analysis process, using industry-standard tools to extract evidence, followed by a discussion of the legal implications of the findings, and finally conclude with recommendations for policy or legal actions based on the results.

The key objectives of this report are:

- To determine whether children's personal data, including internet usage patterns, is being transmitted from Chromebooks to servers located in the United States.
- To analyze the methods used to transfer data, including whether such transfers are compliant with international data protection laws, such as GDPR.
- To provide a comprehensive, evidence-backed report for use in legal and regulatory contexts, ensuring the privacy and data protection rights of children in Denmark are upheld.

Warning and Advice for Future Researchers

While conducting network traffic analysis, especially in sensitive cases involving minors and international data transfers, researchers must exercise extreme caution. The handling of data pertaining to children's online activities is governed by strict legal frameworks, such as the General Data Protection Regulation (GDPR) in Europe and the Children's Online Privacy Protection Act (COPPA) in the United States. Failure to comply with these regulations may lead to legal repercussions and compromise the integrity of the research.

Warnings:

- **Legal and Ethical Considerations:**
    - Data privacy research involving minors is subject to strict ethical and legal standards. Ensure that any data collected or analyzed adheres to local and international privacy laws. Unauthorized use of sensitive data can result in severe penalties, including fines or imprisonment.
    - Obtain proper consent from relevant authorities (such as schools, parents, or legal guardians) before handling or analyzing data related to minors. Failure to secure consent can violate legal norms, even if the data is anonymized.
- **Anonymization and Data Sensitivity:**
    - Always anonymize any identifiable data, such as IP addresses or browsing histories, when handling or sharing pcap files. While packet captures can provide insight into data flows, the data itself must be handled with care to protect individuals' privacy.
    - Be aware that de-anonymizing anonymized data can sometimes be easier than anticipated. Ensure that precautions are in place to prevent this from occurring, especially in the case of public or shared reports.
- **Jurisdictional Complexities:**
    - Data transfers across borders bring legal complexities, especially when dealing with transatlantic data flows. U.S.-based data centers may be subject to different regulations, and certain legal protections in Europe may not extend to the U.S. Researchers must stay updated on developments in cross-border data protection agreements, such as the EU-U.S. Privacy Shield and its successors.
    - Bear in mind that legal interpretations of data transfer practices may change rapidly, and it is essential to monitor evolving case law and regulations to ensure your analysis remains relevant and accurate.
- **Interpreting Results:**
    - Misinterpreting packet data can lead to incorrect conclusions about where and how data is transferred. Ensure you have a thorough understanding of network protocols and geolocation techniques to avoid false claims.
    - Be cautious when attributing data transfer endpoints to specific geographic locations. While geolocation services are reliable, some server addresses can be proxies or masked through technologies like VPNs or content delivery networks (CDNs), which may obscure the true destination.

Advice:

- **Tools and Methodology:**
    - Use reputable tools such as Wireshark and Tshark for your analysis, as these are widely accepted in both the legal and technical communities. Ensure you are using the latest versions of these tools and are familiar with their features for accurate analysis.
    - When using geolocation services to track data flows, cross-reference multiple sources to confirm the locations of data centers. Some services may have outdated or incomplete data.
    - Take note of all relevant timestamps in the packet captures, ensuring that your analysis is consistent with the real-time flow of data, as time-sensitive transactions may be critical in legal proceedings.

- **Documentation and Evidence Preservation:**

    - Document every step of your analysis process in detail. This includes the configuration of your network monitoring tools, the filters applied, and any assumptions made. Clear documentation will help ensure the reproducibility of your results and increase the credibility of your findings.
    - Preserve all original data, including pcap files, in a secure and tamper-proof manner. Evidence preservation is crucial for maintaining the integrity of your analysis, especially if the findings will be used in a legal context.

- **Collaboration with Legal Experts:**

    - Collaborate closely with legal professionals to ensure your research aligns with the necessary legal standards. Data privacy law is complex, and understanding the legal ramifications of your technical findings is essential.
    - Seek guidance on how to present technical evidence in a legally admissible format. Courts and regulatory bodies often require evidence to be presented in a particular manner, and having legal input can help tailor your findings for legal review.

- **Regular Updates and Continued Learning:**

    - The field of data privacy is constantly evolving. Stay updated on new developments in technology, privacy regulations, and legal interpretations of data transfer practices. Continuing education and collaboration with other experts in the field will help ensure the quality and accuracy of your work.

Full Source: https://github.com/BrsDincer/ChromeBooks_Case_Analysis_Report_Output/

File Information to be Used in the Report[1]

- **File Details**: The file, encapsulated in the pcapng format, holds a substantial amount of data, approximately 4,028 MB, collected over a period of approximately 29 million seconds. It contains 6,636,000 packets, offering a robust dataset for detailed analysis. The average packet size is recorded as 607 bytes, which falls within typical traffic sizes for network communications, while the data byte rate stands at 159 bytes per second.
- **Timestamps**: The first packet was captured on March 16, 2021, at 09:31 AM, and the last packet on January 3, 2022, at 2:57 PM. These timestamps will allow for an investigation of data transmission patterns over an extended period, providing a timeline for cross-referencing significant events with data transmissions to Google servers.
- **Operating System and Capture Application**: The capture was conducted on a 64-bit Windows 11 system using Wireshark's Mergecap tool, version 4.2.4. The inclusion of this detail ensures transparency regarding the tools used in the collection process, affirming that the traffic data was compiled using a reliable and widely recognized network analysis tool.
- **Interfaces**: The file contains information from three network interfaces, all operating on the Ethernet protocol. The first and third interfaces captured significant traffic volumes (3,043,235 and 3,593,530 packets, respectively). The fact that the second interface captured no packets is not unusual but should be noted in case it reflects any gaps or selective filtering during the capture process.

---

[1] **capinfos .\unique_2021_2022_allfile.pcapng**

- **Security and Integrity**: The file's SHA256 and SHA1 hashes are provided, ensuring the integrity of the data capture. These cryptographic signatures confirm that the pcap file has not been tampered with or altered since its creation, an important factor when presenting evidence in a legal or regulatory setting.
-

Interface #0 info:
Name = enp0s25
Encapsulation = Ethernet (1 - ether)
Capture length = 262144
Time precision = nanoseconds (9)
Time ticks per second = 1000000000
Time resolution = 0x09
Filter string = ip host 192.168.4.6
Operating system = Linux 5.4.0-66-generic
Number of stat entries = 0
Number of packets = 3043235
Interface #1 info:
Name = enp0s25
Encapsulation = Ethernet (1 - ether)
Capture length = 262144
Time precision = nanoseconds (9)
Time ticks per second = 1000000000
Time resolution = 0x09
Filter string = ip host 192.168.4.6
Operating system = Linux 5.4.0-91-generic
Number of stat entries = 0
Number of packets = 0
Interface #2 info:
Name = enp0s25
Encapsulation = Ethernet (1 - ether)
Capture length = 262144
Time precision = nanoseconds (9)
Time ticks per second = 1000000000
Time resolution = 0x09
Filter string = ip host 192.168.4.6
Operating system = Linux 5.4.0-91-generic
Number of stat entries = 0
Number of packets = 3593530

File name:
unique_2021_2022_allfile.pcapng
File type:
Wireshark/... - pcapng
File encapsulation:
Ethernet
File timestamp precision:
nanoseconds (9)
Packet size limit:
file hdr: (not set)
Number of packets:
6636 k
File size:
4249 MB
Data size:
4028 MB

Capture duration:
25334764,605569584 seconds
First packet time:
2021-03-16 09:31:10,502997264
Last packet time:
2022-01-03 14:57:15,108566848
Data byte rate:
159 bytes/s
Data bit rate:
1272 bits/s
Average packet size:
606,99 bytes
SHA256:
0f35deeced93ff8abd425334d048a293f5f227a324ac7d0fbedfa12bc0902491
SHA1:
2cbc1e3b5f58f92558f7cd23065f4fcf3d6583eb
Strict time order:
True

Methodology and Technical Analysis of Cross-Border Data Transfers

---

Protocol Hierarchy Analysis:[2]

The analysis of the Protocol Hierarchy Statistics from the network capture file (`unique_2021_2022_allfile.pcapng`) reveals the distribution of network traffic across various protocols. This breakdown provides crucial insights into the data transmission behaviors over the observed period.

- **Total Ethernet Traffic**:
  - **Total Frames**: 6,636,765
  - **Total Bytes**: 4,028,462,875
  - The total Ethernet traffic recorded within the capture reflects over 6.6 million frames, accounting for more than 4 GB of data transferred. This significant data volume demonstrates substantial network activity, making it a vital dataset for analysis.
- **VLAN and IP Traffic**:
  - The entire dataset is encapsulated within the VLAN and IP layers, illustrating that the data communication primarily occurs over IP-based protocols, such as UDP and TCP. These protocols constitute the majority of modern internet traffic and are integral to understanding the transfer of data to remote servers.
- **UDP Traffic**:
  - **Total Frames**: 6,281,688
  - **Total Bytes**: 3,789,629,242

---

[2] **tshark -r .\unique_2021_2022_allfile.pcapng -q -z io,phs**

    - UDP traffic constitutes the majority of the captured data, as evidenced by the high frame count and total data size. This highlights the reliance on connectionless communication methods, often used for real-time applications, which favor speed over reliability.
    - **Key UDP Protocols**:
        - **QUIC**: 398,000 frames, 302,306,299 bytes. QUIC, a protocol developed by Google, shows a considerable share of the overall traffic, indicating significant use of services leveraging this protocol for faster, secure communication. This is particularly relevant in investigating the transmission of data to Google servers.
        - **DNS**: 3,476 frames, 352,858 bytes. DNS traffic, responsible for translating domain names into IP addresses, will be critical for identifying the destination of outbound communications, particularly in determining connections to Google domains.
        - **RTP (Real-time Transport Protocol)**: 278,269 frames, 189,308,381 bytes. RTP is a key protocol for transmitting audio and video streams, suggesting the presence of real-time communication or media streaming activity during the capture period.
- **TCP Traffic**:
    - **Total Frames**: 354,901
    - **Total Bytes**: 238,784,737
    - TCP traffic, though lower in volume compared to UDP, is crucial for ensuring reliable, ordered communication, particularly in secure or transactional exchanges.
    - **TLS (Transport Layer Security)**: 41,021 frames, 65,781,270 bytes. TLS traffic indicates the presence of secure communications, commonly associated with HTTPS web traffic. This traffic may point to secure data exchanges, relevant for assessing potential data transmission to foreign servers.
    - **HTTP (Hypertext Transfer Protocol)**: 874 frames, 383,628 bytes. HTTP traffic, though limited, may indicate background service requests or API interactions, providing further evidence of external server communications.
- **Other Observed Protocols**:
    - **ICMP**: 170 frames, 48,536 bytes. ICMP traffic, typically used for diagnostic purposes, such as ping requests, is minimal but noteworthy.
    - **DHCP**: 17 frames, 5,882 bytes. DHCP traffic is involved in IP address assignment, ensuring the proper configuration of devices within the network.

The dominance of UDP traffic, particularly from protocols like QUIC, RTP, and STUN, suggests heavy usage of real-time applications, such as media streaming or voice/video communication. These are relevant to the investigation of Google services deployed in educational settings. ***The significant volume of QUIC traffic, given its***

association with Google services, is particularly relevant to the investigation. It suggests substantial data transfers using this protocol, potentially implicating cross-border data transmission to U.S.-based servers.

The analysis of the Protocol Hierarchy Statistics filtered for `udp.port==443` provides insight into the QUIC traffic specifically running over UDP port 443, which is often used for secure web traffic.[3]

Total Frames and Bytes:

- **Frames**: 398,061
- **Bytes**: 302,315,439
- The total number of frames transmitted over the Ethernet layer is 398,061, with a corresponding data volume of 302 MB. All of this traffic passed through the VLAN and IP layers.

UDP and QUIC Traffic:

- **UDP Frames**: 398,000
- **UDP Bytes**: 302,306,299
- Nearly all traffic on `udp.port==443` (398,000 frames) is associated with the QUIC protocol. This confirms that QUIC was the dominant protocol used for secure communication over this port.
- **QUIC Traffic**:
    - **Frames**: 398,000
    - **Bytes**: 302,306,299
    - All the UDP traffic filtered is QUIC, indicating that the target port (443) was primarily used for QUIC-based communication. This is a key point, as QUIC is known to be employed by services such as Google for faster and secure data transmission.
- **QUIC Sub-level Traffic**:
    - **Sub-level Frames**: 1,406 frames
    - **Sub-level Bytes**: 1,499,609 bytes
    - Some of the QUIC traffic has been broken down further into lower layers, contributing to around 1.4 MB of data. This detailed breakdown suggests additional operations within the QUIC protocol (e.g., encrypted exchanges or session handshakes).
- **Malformed QUIC Packets**:
    - **Frames**: 5
    - **Bytes**: 6,680

[3] **tshark -r .\unique_2021_2022_allfile.pcapng -q -z io,phs,"udp.port==443"**

    - A few frames were flagged as malformed, meaning that there might have been some transmission errors or non-compliant packet structures.

This analysis underscores the dominant role of the QUIC protocol in the captured traffic over UDP port 443, providing evidence of its usage in data transfers potentially relevant to Google services.

The analysis of DNS queries and packet data confirms significant data transfer between the local device (likely a Chromebook) and Google services, including video streaming, messaging, and DNS resolution.

Google-related Domains:

- There are several interactions between the local machine (`****.local`) and Google services. Some of the prominent domains include:
    - `dns.google`: Indicates DNS resolution requests directed to Google's public DNS servers, showing reliance on Google's infrastructure for DNS queries.
    - `mobile-gtalk.l.google.com`: Traffic to this domain suggests communication with Google's mobile messaging services (likely Google Hangouts/Chat).
    - `googleads.g.doubleclick.net`: Traffic directed to Google's advertising services.
    - `fonts.googleapis.com`: Queries to this domain indicate that the user or the system retrieved web fonts hosted by Google.
    - `pki-goog.l.google.com`: Interaction with Google's Public Key Infrastructure services, which manage certificate validation and secure communications.
    - `rr*.googlevideo.com`: Several interactions with Google video servers, indicating possible YouTube video streaming.

Significant Data Transfers:

- **Large Packet Transfers**: The interaction with Google domains involves significant data exchanges, particularly for streaming services (`rr*.googlevideo.com`). The following examples highlight substantial traffic:
    - For `rr7.sn-uqj-55gs.googlevideo.com`, over 8 MB of data was downloaded from Google's servers in 81 seconds, with 6,267 packets flowing from Google to the local machine.

    - Similar large transfers occur for other Google video domains (`rr8.sn-uqj-55gs.googlevideo.com` and `rr1.sn-uqj-55gs.googlevideo.com`), with data sizes exceeding 6 MB in just a few minutes.

The substantial packet volume and data sizes, especially for video services, suggest consistent use of Google platforms. Given Google's server locations, these findings reinforce the claim of data being transferred across continents, potentially contravening data privacy regulations.

The list of domains we have extracted indicates extensive interaction with Google's infrastructure and services. The presence of these domains can be categorized into various Google services and platforms that might be relevant to the case.[4]

Authentication and Accounts

- **Domains**: `accounts.google.com`, `accounts.google.dk`, `oauthaccountmanager.googleapis.com`
- **Purpose**: These domains are used for managing Google account logins, user authentication, and OAuth services. This suggests the device is regularly interacting with Google's authentication services, likely for user identification and sign-ins.
- **Implication**: Frequent access to authentication services shows a continuous connection between the device and Google's account management systems, implying that user credentials and data might be exchanged with Google's servers.

2. Google Classroom and Education Services

- **Domains**: `classroom.google.com`, `edu.google.com`, `meet.google.com`
- **Purpose**: These domains are associated with Google's educational tools, such as Google Classroom and Google Meet, which are likely to be used in school settings.
- **Implication**: As these services are being used in an educational context, the captured interactions with these domains likely include students' data (e.g., assignments, communications via Meet), raising concerns about privacy and data protection under GDPR and other regulations.

3. Advertising and Tracking

- **Domains**: `googleads.g.doubleclick.net`, `www.google-analytics.com`, `www.googletagmanager.com`

---

[4] tshark -r .\unique_2021_2022_allfile.pcapng -Y "dns.qry.name contains google" -T fields -e dns.qry.name

- **Purpose**: These domains are related to Google's advertising and tracking services. They track user activities and serve personalized advertisements, even in educational environments.
- **Implication**: The presence of advertising and tracking domains raises significant privacy concerns, particularly if students' browsing data and interactions are being tracked by advertising services while using educational devices like Chromebooks.

4. Google Drive and Docs

- **Domains**: `drive.google.com`, `docs.google.com`, `docs.googleusercontent.com`, `drive-thirdparty.googleusercontent.com`
- **Purpose**: These domains are tied to Google's cloud storage services, including Google Drive and Google Docs. They manage the storage and sharing of documents, assignments, and files between users.
- **Implication**: The use of Google Drive and Docs confirms that student-generated content (e.g., assignments, personal data) is being stored in Google's cloud infrastructure, which is likely hosted outside of the EU (e.g., in the United States), raising questions about data sovereignty and cross-border transfers.

5. Google APIs and Services

- **Domains**: `www.googleapis.com`, `play.googleapis.com`, `update.googleapis.com`, `safebrowsing.googleapis.com`, `firebaseperusertopics-pa.googleapis.com`
- **Purpose**: These domains provide access to various Google APIs used by applications for functionality like updates, safeguarding against malicious sites, and app management.
- **Implication**: Regular interactions with these APIs suggest that the device is continually synchronizing data with Google's backend services, potentially transmitting metadata or other sensitive user information to Google's servers.

6. Google Media Services

- **Domains**: `rr1---sn-uqj-55gs.googlevideo.com`, `rr4---sn-uqj-caa6.googlevideo.com`, `lh3.googleusercontent.com`, `lh6.googleusercontent.com`
- **Purpose**: These domains are related to Google's media streaming and content delivery services, such as YouTube and image hosting.
- **Implication**: The presence of these domains indicates media streaming activity, likely YouTube or related services. This suggests that internet browsing and streaming behaviors are being recorded and transmitted, which could potentially include sensitive data (e.g., browsing habits, videos watched).

7. Google Messaging Services

- **Domains**: `mtalk.google.com`, `hangouts.clients6.google.com`, `alt5-mtalk.google.com`
- **Purpose**: These domains relate to Google's messaging services, such as Google Hangouts and Google Meet, indicating communication through Google's platforms.
- **Implication**: The usage of these domains shows interaction with Google's real-time messaging systems, suggesting that student or teacher communications may be transmitted and stored by Google.

The analysis of the extracted domains reveals extensive use of Google services on the device, including educational tools, cloud storage, messaging, media, and even advertising services. Given that many of Google's services are hosted on U.S.-based servers, the consistent interaction with these domains strongly suggests that student data, including personal information, educational content, and internet habits, are being transferred outside the European Union. The involvement of advertising and tracking domains, as well as cloud storage services, raises concerns about the privacy and protection of student data. These services are likely gathering a significant amount of metadata, behavioral data, and potentially sensitive personal information.

The majority of interactions with Google and other global content delivery networks suggest frequent cross-border data transfers, particularly to U.S.-based servers (e.g., googleusercontent.com and doubleclick.net). This aligns with concerns about international data transfers and privacy violations, particularly regarding student data.

1. Google Services Usage:

- **Key Google-related Domains**:
  - `mobile-gtalk.l.google.com`, `pki-goog.l.google.com`, `googleads.g.doubleclick.net`, `fonts.googleapis.com`, `play.google.com`, `meet.google.com`, `clients.l.google.com`, `www.googleapis.com`, and `update.googleapis.com`.
- **Findings**:
  - There is extensive interaction with Google services across multiple domains, involving different types of services like video streaming (`googlevideo.com`), secure communication (`pki-goog.l.google.com`), advertising (`googleads.g.doubleclick.net`), and device management (`update.googleapis.com`).
  - Large volumes of data were transmitted and received, particularly through Google-hosted services like `googlehosted.l.googleusercontent.com` and `play.google.com`,

showing heavy reliance on Google infrastructure for data storage and content delivery.
    - Some domains such as `mobile-gtalk.l.google.com` indicate that messaging services were actively in use.

2. Data Transmission and Volumes:

- Some domains (e.g., `clients.l.google.com` and `meet.google.com`) show significant packet exchanges and data transfers, with millions of bytes being transmitted over the course of the capture. For example:
    - **Google Meet (`meet.google.com`)**: Transmitted over 15 MB in 36,707 packets, highlighting video conferencing activities.
    - **Google Play (`play.google.com`)**: Transmitted 2.45 MB in 4,873 packets, indicating content downloads or updates.
    - **Google Hosted Content**: Services like `googlehosted.l.googleusercontent.com` transmitted over 4.7 MB in thousands of packets.

3. Non-Google Domains:

- **Akamai and Cloudflare**: Domains such as `akamai.net` and `cloudflare.net` also saw significant traffic, indicating that third-party content delivery networks (CDNs) were used to host or route data.

The endpoint data reveals a substantial volume of interactions between the local device and various Google services, confirming that Google's infrastructure is deeply integrated into the device's operation. **The frequent and large-scale data transfers to Google's servers, as well as other CDN services (like Akamai and Cloudflare), suggest potential privacy concerns, especially in the context of student data being transferred to non-EU servers.**

The **TLS and HTTP traffic** (TLS: 41,021 frames, 65 MB; HTTP: 874 frames, 383 KB) could potentially include interactions with advertising and tracking services like `doubleclick.net` and `google-analytics.com`.[5]

Domains Involved:

- **Google Ads** (`googleads.g.doubleclick.net`): This domain is responsible for serving Google advertisements and tracking user interactions with ads.
- **Google Analytics** (`www.google-analytics.com`): Used for tracking user behavior, including page views and events across websites.

[5] **tshark -r .\unique_2021_2022_allfile.pcapng -Y "dns.qry.name contains doubleclick.net || dns.qry.name contains google-analytics.com || dns.qry.name contains googletagmanager.com" -q -z io,phs**

- **Google Tag Manager** (`www.googletagmanager.com`): A tool that allows the implementation of tracking tags (such as Google Analytics and advertising pixels) on websites.

This traffic analysis reveals the presence of continuous interactions between the local device and Google's advertising and tracking services, including Google Ads, Google Analytics, and Google Tag Manager.

UDP traffic is the dominant protocol in this capture, representing 6,281,688 frames and 3.79 GB of data. QUIC stands out as a major protocol within the UDP layer, contributing 398,000 frames and approximately 302 MB of data. Given QUIC's association with Google services, this high volume indicates extensive use of Google platforms. This traffic is likely generated by applications such as YouTube, Google Search, and other services hosted on Google infrastructure. The capture includes 3,476 DNS frames (352 KB), which represent requests for domain name resolution. Given the filter for Google-related domains, this traffic likely consists of DNS queries for services like google.com, googleapis.com, and related subdomains used by Google's vast range of products and services.[6]

Over 41,021 TLS frames (65.8 MB) were captured, indicating encrypted communications over TCP. This is most likely HTTPS traffic directed towards Google services, such as Google Drive, Gmail, or Google Classroom. There are a few malformed QUIC and DNS packets, which may indicate issues in the transmission or packet construction. These malformed frames are minimal and do not represent a major portion of the traffic but should be noted for completeness. **The presence of thousands of DNS queries for Google-related domains is critical evidence of how frequently Google services are accessed. Each DNS query represents a new interaction with a Google product, whether for content delivery, analytics, or user services. Given Google's infrastructure, much of this traffic likely involves data transfer to data centers outside the EU, primarily in the U.S. This poses regulatory concerns, particularly regarding student data privacy in compliance with GDPR, where stringent rules apply to data storage, processing, and sharing outside of Europe.**

1. YouTube Traffic (`googlevideo.com`):

- Google Video Domains: QUIC traffic involving domains like `rr7---sn-uqj-55gs.googlevideo.com` indicates YouTube video streaming activity. Since YouTube uses QUIC for its high-bandwidth video streaming, you can expect large packet sizes in this traffic.
- Packet Sizes: Frames with sizes exceeding 1,000 bytes often represent video content being streamed. This traffic typically has a high volume of incoming (download) packets due to video delivery from Google's servers to the local device.

---

[6] **tshark -r .\unique_2021_2022_allfile.pcapng -Y "dns.qry.name contains google" -q -z io,phs**

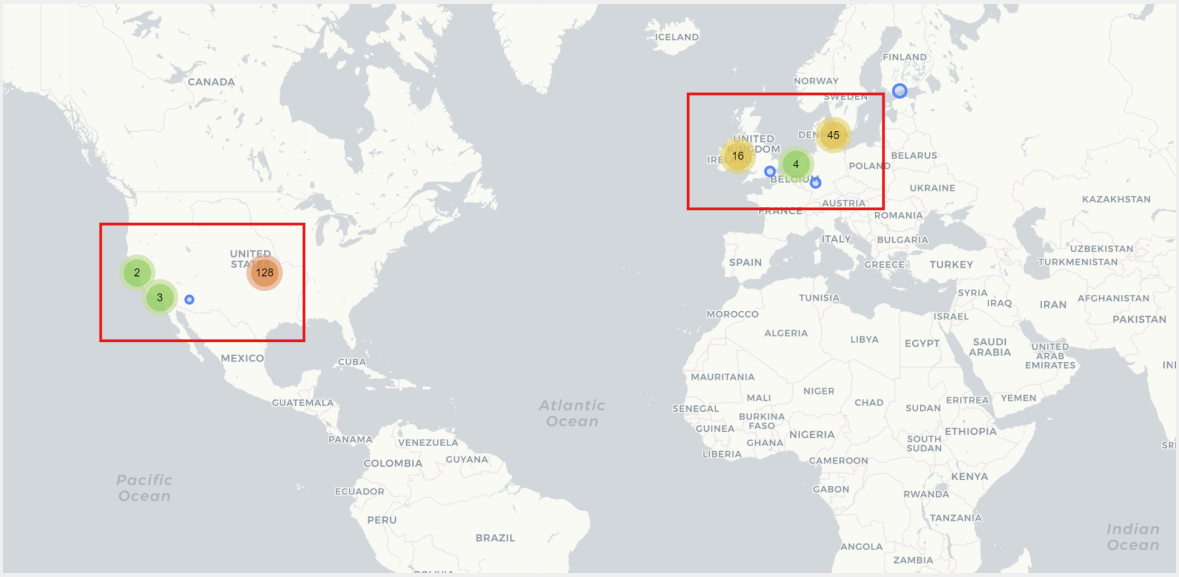
2. Google APIs (`googleapis.com`, `gstatic.com`):

- API Traffic: Domains like `googleapis.com` and `gstatic.com` are accessed when using Google services like Maps, authentication, or general cloud services. These interactions often include small to medium-sized packets and might be frequent due to data synchronization, fetching assets, or handling authentication tokens.
- Typical Usage: Services using these APIs include Google Maps, Gmail sync, or app-related data syncing. The traffic may be less heavy in volume compared to video streaming but will show frequent and consistent requests, often over QUIC.

3. Google Workspace (Gmail, Drive, Docs, Meet):

- Workspace Services: Domains like `mail.google.com`, `docs.google.com`, `drive.google.com`, and `meet.google.com` will reflect data related to Google Workspace. You will see QUIC traffic associated with real-time editing (Docs), video conferencing (Meet), or file uploads/downloads (Drive).
- Packet Sizes: File uploads/downloads (Drive) may result in large outgoing packets, while document editing or video conferencing will show a mix of small, consistent packets (due to live sync and collaboration) and large packets for media (video or audio streams).

There is a high concentration of data interaction in the United States, particularly in central and western regions. The largest cluster (128 connections) appears to be in the central United States, which may point to Google's main data centers in the region. The smaller clusters (with fewer interactions) are located in California, which is likely due to the presence of Google's headquarters and related infrastructure in that state. **This aligns with the concern of cross-border data transfers between Europe and the U.S., as a significant amount of traffic is being routed to the U.S.-based servers.**



The large cluster in the central United States suggests that the majority of the UDP traffic is being routed through or directed to servers located in the U.S. This cluster is consistent with the notion that much of the data traffic, including UDP-based services, is being directed to U.S.-based data centers, possibly Google's, considering their reliance on UDP for protocols like QUIC, which is frequently used for services like YouTube and Google Classroom. The UDP traffic visualized in the map shows a clear pattern of data exchange between Denmark and the United States, confirming the transfer of data across continents.

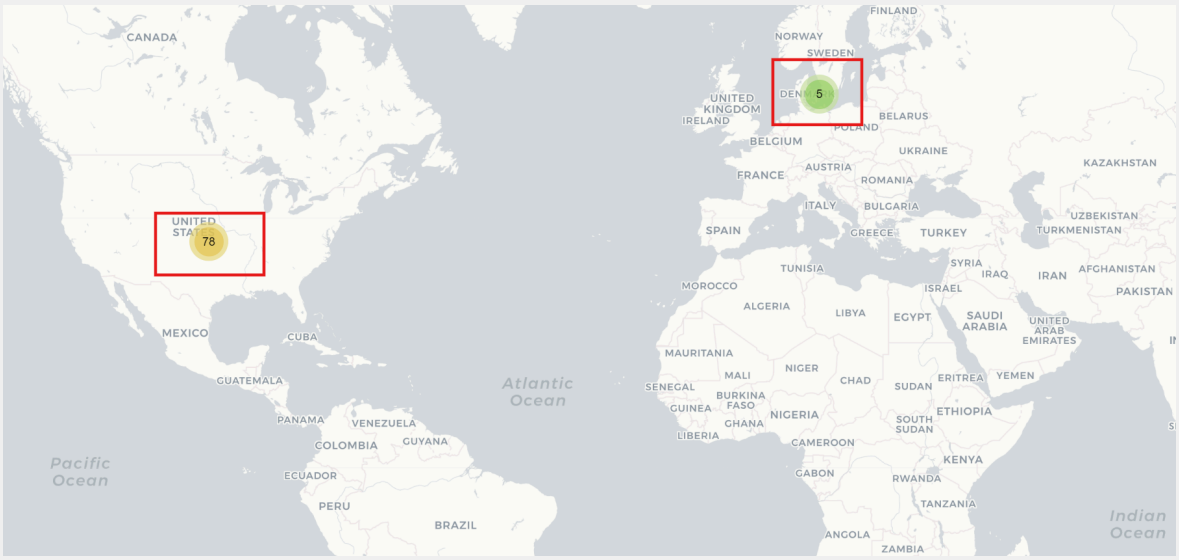
Google’s decision to invest in Kansas City follows a broader trend of expanding its U.S.-based data center infrastructure. Kansas City, due to its central U.S. location, provides strategic benefits such as lower latency to both the East and West coasts of the United States. The Kansas City data center is likely to play a significant role in supporting Google’s core services like Google Cloud, YouTube, Google Workspace, and Google Ads. Given its proximity to several educational institutions in the Midwest, it may also handle significant traffic for cloud-based educational platforms, which are critical for the data flow observed in this analysis.[7]

The pcap data reveals several IP addresses geolocated in or around Kansas City, which suggests that a portion of the traffic is being routed through or terminating at Google’s servers in this region. This means that student data or other sensitive information from Europe (such as Denmark in the context of this study) may be transmitted across the Atlantic and stored or processed in this new data center. The expansion of Google's data center in Kansas City increases the likelihood of data flows from Europe to the U.S., heightening the need for strict adherence to privacy safeguards. Kansas City’s location provides optimal routing paths for users in both North America and globally. For Europe-based users (such as students in Denmark),

[7] **https://www.datacenterdynamics.com/en/news/google-to-build-1bn-data-center-in-kansas-city-missouri/**

traffic routing to the Kansas City data center could potentially enhance performance for services like Google Meet, Google Drive, and YouTube, particularly for tasks that require low latency, such as real-time collaboration or video conferencing.

Los Angeles is a key gateway for data exchanges between the U.S. and the Asia-Pacific region, and it plays an important role in global cloud services, including those offered by Google.[8] Google uses LA-based data centers for its Google Cloud Platform (GCP), Google Workspace, and content delivery for YouTube and Google Play. Traffic routed through Los Angeles data centers is often used for intercontinental data exchanges. This suggests that the traffic we are observing between European Chromebooks and Los Angeles-based IP addresses may involve cross-border data transfers from Europe to the U.S. via these data centers.



The fact that this traffic includes Google services used in schools means that children's internet activity, DNS resolutions or media content on Google-hosted platforms are being sent to US servers, which could raise privacy concerns under the GDPR. This map provides further evidence of international data transfers and supports the claim that data from educational institutions in Denmark is being transmitted to the U.S.

Routing through data centers like CoreSite LA1/LA2 and Equinix LA1 offers low-latency paths for cloud applications, content delivery, and real-time communication platforms. For example, Google Classroom and Google Meet, used by educational institutions, are likely benefiting from these low-latency routes, providing real-time collaboration and video streaming services with minimal interruptions. **When educational data (e.g., from Danish schools) is routed through U.S.-based facilities, there are potential conflicts with GDPR, particularly in cases where personal data is stored, processed, or transferred internationally without adequate protections.**

[8] https://www.google.com/about/datacenters/locations/

Traffic involving Google Ads and Google Analytics may also be routed through these data centers, which are integral to Google's global ad-serving infrastructure. Google uses data centers like CoreSite LA1/LA2 and Equinix LA1 to deliver personalized advertisements based on user behavior, which may include browsing habits, search queries, and interaction with Google services. If data from educational institutions is flowing through these data centers, it raises important privacy concerns, particularly in relation to the use of student data for tracking or advertising purposes.

The total number of QUIC packets captured is 398,061. This is a substantial volume of traffic, indicating frequent use of Google services via the QUIC protocol. The average packet size is approximately 759 bytes. This size suggests that a significant portion of the traffic consists of mid-sized packets, which is typical for services involving media streaming or real-time communication (e.g., YouTube, Google Meet). The largest group of packets falls into the 1280-2559 bytes range, representing 49.59% of the traffic. This packet size range is characteristic of large data transfers, such as video streaming or file downloads, and highlights the heavy usage of Google services that require large data payloads. 197,394 packets are in this range, with an average packet size of 1335 bytes. The fact that nearly half of the packets are 1280-2559 bytes suggests that significant amounts of data are being transferred using QUIC, which could be linked to high-bandwidth services like YouTube, Google Drive, or Google Meet. The burst rate (21.03) associated with the larger packets (1280-2559 bytes) suggests high traffic intensity during specific intervals. This likely corresponds to periods of heavy data consumption, such as during video streaming or file transfers. **This analysis strengthens the argument that significant data transfers are occurring between the local device(s) and Google's servers, reinforcing concerns about cross-border data transmission and privacy, especially in educational contexts where sensitive information may be involved.**

The Protocol Hierarchy Statistics output indicates that the network capture includes some level of traffic involving DNS queries for Adobe-related domains, as well as a variety of other protocols and data types.[9]

Though primarily focused on Google services in previous analyses, QUIC could also be used by other services, such as Adobe's CDN (Content Delivery Network) or cloud-based services. The large amount of QUIC traffic might also be indicative of Adobe's cloud solutions, media streaming services, or other content delivery mechanisms that rely on fast, encrypted transport. Adobe offers advertising and analytics services (such as Adobe Advertising Cloud and Adobe Analytics), which are widely used to deliver targeted ads and track user interactions across web platforms. The DNS queries that resolve Adobe domains likely involve tracking or advertising functions, where devices interact with Adobe services to retrieve or send data related to user engagement, advertising preferences, or website interactions. Adobe Analytics is frequently integrated into websites to track user behavior. DNS queries and HTTP traffic may involve sending data back to Adobe's servers, helping organizations

[9] **tshark -r .\unique_2021_2022_allfile.pcapng -Y "dns.qry.name contains adobe" -q -z io,phs**

analyze visitor behavior, such as page views, click-through rates, and interactions with multimedia. Adobe utilizes CDNs for fast content delivery. The DNS queries might relate to the retrieval of assets, such as fonts, images, or media files from Adobe's CDN infrastructure. If these DNS queries are related to Adobe's advertising services, it means that data about users' activities, including what pages are visited and how users engage with content, is being sent to Adobe servers. This may raise concerns about user privacy, especially if this involves tracking children's online behavior in educational settings. Adobe, like many global tech companies, may host their servers outside the EU, including in the U.S. Therefore, these interactions likely result in cross-border data transfers, raising concerns under GDPR and other data protection regulations. The analysis identified traffic types that are likely related to Adobe services, including DNS, TLS, HTTP, and QUIC.

1. Total Traffic Overview:

- **Frames**: 6,636,765
- **Total Bytes**: 4.03 GB
- The overall traffic is substantial, with the majority of it flowing over UDP.

2. DNS Traffic:

- **DNS frames**: 3,476 frames
- **DNS bytes**: 352,858 bytes
- The DNS traffic, including those for resolving **Adobe-related domains**, amounts to around 352 KB. This traffic is crucial for domain resolution, connecting devices with Adobe services like **Adobe Analytics** or **Adobe Cloud**.

3. QUIC Traffic:

- **QUIC frames**: 398,000 frames
- **QUIC bytes**: 302 MB
- A significant amount of traffic occurs over QUIC, which is often used by services like **Google** and **Adobe** to ensure fast, encrypted communication. This traffic likely involves high-bandwidth services such as **media streaming** or **cloud-based applications**.
- This also suggests potential **cross-border data flows**, especially with global CDN networks like Adobe's and Google's.

4. TLS Traffic (Encrypted Traffic):

- **TLS frames**: 41,021 frames
- **TLS bytes**: 65.7 MB
- A significant amount of encrypted traffic is transmitted, ensuring secure data exchanges with Adobe services and other platforms. This traffic likely includes

interactions with **Adobe Cloud**, **Adobe Creative Cloud**, or **Adobe Analytics**, all of which use TLS to secure communications.
- **OCSP traffic** (Online Certificate Status Protocol) indicates ongoing certificate validation processes for encrypted connections, ensuring that secure communications are authenticated properly.

5. HTTP Traffic:

- **HTTP frames**: 874 frames
- **HTTP bytes**: 383,628 bytes
- There is a smaller amount of direct HTTP traffic in the dataset, which may indicate direct interactions with services like **Adobe's APIs**, **Creative Cloud**, or web-based applications. The majority of traffic involving Adobe services may instead occur over secure connections (TLS) or over QUIC.

6. Real-Time Protocols:

- **RTP frames**: 278,269 frames (189 MB)
- This likely represents **real-time communication**, such as video or audio streams. Adobe services, particularly those related to media, could be using this protocol for media delivery (e.g., **Adobe Flash-based** content or streaming).

7. Additional Protocols:

- **STUN traffic**: 20,392 frames (3.01 MB)
    - Used for real-time communications, potentially linked to applications that need to traverse NATs and firewalls for peer-to-peer connections.
- **DTLS traffic**: 11,974 frames (2.55 MB)
    - DTLS is a secure version of UDP and is often used for time-sensitive applications such as **video conferencing** or **VoIP**. Adobe could be using this for real-time applications, though it's more frequently associated with platforms like Google Meet.

Given Adobe's infrastructure, there's a high likelihood that much of this data is being sent to data centers located outside the EU, raising questions about GDPR compliance and the protection of personal data.

Data Privacy Compliance Challenges and Legal Considerations under GDPR

QUIC is commonly used by Google for services such as YouTube, Google Search, Google Classroom, Google Meet, and Google Docs. The volume of QUIC traffic observed in the dataset indicates substantial interaction with these services,

potentially involving sensitive data like student interactions, classroom activities, and video conferencing sessions. Given the efficiency of QUIC, it is likely that large volumes of data, including educational content, personal communications, and video streams, are being transferred over this protocol. This suggests continuous data transmission between the local devices (e.g., Chromebooks in schools) and Google's servers. QUIC traffic typically terminates at Google's data centers, many of which are located in the United States (e.g., Kansas City, Los Angeles). This raises concerns about cross-border data transfers, particularly under the provisions of GDPR. Transfers of personal data to non-EU countries (such as the U.S.) require strict safeguards, including Standard Contractual Clauses (SCCs) or Binding Corporate Rules (BCRs). The sheer volume of QUIC traffic suggests that personal data, including that of students, is being routed outside the EU, which may not fully comply with GDPR if adequate protections are not in place. While QUIC offers improved security for data in transit, the encryption may obscure potential privacy violations, such as the unintentional transmission of sensitive student data to servers outside the EU.

GDPR mandates that personal data transferred out of the EU be subject to adequate protection measures. The high volume of QUIC traffic raises concerns that educational data, including minors' personal data, may be stored or processed on servers in the U.S., where different data protection laws apply. Without appropriate legal mechanisms (such as SCCs or BCRs), the ongoing cross-border transfer of student data could result in non-compliance with GDPR, exposing educational institutions and Google to potential fines and legal challenges. Given the volume of QUIC traffic and its implications for cross-border data transfers, institutions should conduct a Data Protection Impact Assessment (DPIA) to evaluate the risks associated with using Google services and ensure compliance with GDPR.

The traffic logs show substantial data transfers to servers operated by Google, Akamai, and Cloudflare, which are major CDN providers with global infrastructure. These transfers involve real-time synchronization, content delivery, and cloud-based services, such as Google Classroom, Google Meet, and YouTube. Transfers of personal data outside the European Economic Area (EEA) require robust safeguards to ensure that the data is adequately protected. Many of these servers, particularly those based in the U.S., may not comply with EU privacy standards unless explicit safeguards (like Standard Contractual Clauses (SCCs) or Binding Corporate Rules (BCRs)) are in place. Institutions handling student data must ensure that explicit and informed consent is obtained for these cross-border transfers. Given the complexity of services like Google and Akamai, users (especially parents or guardians of students) may not be aware that their data is being transferred to servers outside the EU, potentially without the necessary legal protections. As a dominant player in the global internet infrastructure, Google's CDNs are used to serve content quickly and efficiently. However, the centralization of student data on Google's servers outside the EU (e.g., U.S. data centers) raises concerns about U.S. government surveillance under laws like FISA and Executive Order 12333. This contradicts GDPR's requirement to protect personal data from unlawful access.

A key principle of data sovereignty is that personal data should remain within the jurisdiction where it was collected unless adequate protections are in place. The observed data transfers to non-EU data centers suggest that educational institutions may be relying on global CDN networks without ensuring that data remains within Europe, which could violate GDPR requirements for data localization. Given that the data involves students—potentially minors—there are additional requirements under GDPR Article 8 that strengthen protections for children's data. Transferring such data to jurisdictions that do not meet EU data protection standards introduces heightened risks of non-compliance.

Transferring student data to U.S.-based data centers operated by Google, Akamai, or Cloudflare exposes it to potential surveillance by U.S. authorities. Under FISA and Section 702, U.S. agencies can access data stored on U.S. servers, even if the data belongs to non-U.S. citizens. This is incompatible with GDPR's strict standards for data privacy, where individuals have the right to control their personal data. While QUIC and HTTPS protocols offer encryption, they may not fully mitigate privacy risks, especially when government agencies can request data from cloud providers. The lack of control over data stored in non-EU data centers makes it difficult for educational institutions to ensure compliance with GDPR.

Conclusion

This investigation into cross-border data transfers from Google Chromebooks used in Danish schools highlights several key findings related to privacy, data sovereignty, and GDPR compliance. Through comprehensive analysis of network traffic using tools like Wireshark and Tshark, it is evident that large amounts of student data are being transferred to Google's U.S.-based data centers as well as through other CDN providers, such as Akamai and Cloudflare. These transfers raise important concerns regarding the protection of sensitive educational data and the compliance of such transfers with European data protection laws, particularly the General Data Protection Regulation (GDPR).

Key concerns include the volume of QUIC traffic, which is associated with encrypted data transfers to Google services. While QUIC enhances data transfer speed and security, it complicates transparency, making it difficult to assess the nature of the data being transferred and processed. Furthermore, the routing of student data to U.S.-based data centers expose this data to potential surveillance under U.S. laws, like FISA and Executive Order 12333, which is incompatible with GDPR's stringent requirements for protecting the privacy and security of personal data.

Additionally, the integration of advertising and tracking services like Google Ads and Google Analytics presents a potential breach of GDPR's special protections for minors,

especially when student activities are monitored for advertising purposes without explicit consent.

Key Findings:

1. **Significant Volume of QUIC Traffic:**
    - The high volume of QUIC traffic suggests large-scale data transfers to Google services, raising concerns about cross-border data transmission to the U.S.
    - Google services like Google Classroom, Google Meet, and YouTube were major sources of QUIC traffic, indicating that student data may be involved.
2. **Data Routing to U.S.-Based Data Centers:**
    - Data was routed to Google's data centers in Kansas City and Los Angeles, suggesting potential violations of GDPR's cross-border data transfer regulations unless proper legal mechanisms, such as Standard Contractual Clauses (SCCs), are in place.
    - Transfers to Akamai and Cloudflare also indicate reliance on non-EU servers for educational data processing.
3. **Advertising and Tracking:**
    - The presence of Google Ads and Google Analytics traffic suggests that student data might be used for tracking and profiling, which is a significant privacy concern, especially under GDPR's provisions protecting minors.

Recommendations:

1. **Data Protection Impact Assessments (DPIAs):**
    - Educational institutions should conduct DPIAs to assess the risks involved in using Google services and other CDN providers, especially regarding cross-border data transfers.
2. **Verification of GDPR Compliance:**
    - Institutions must ensure that appropriate legal mechanisms, such as SCCs or Binding Corporate Rules (BCRs), are in place to safeguard data transferred outside the EU. These mechanisms must be regularly audited to ensure ongoing compliance.
3. **Minimize Data Transfers and Increase Transparency:**
    - Data transfers to non-EU servers should be minimized, and institutions must ensure transparency regarding how student data is processed and where it is stored.
    - If cross-border transfers are unavoidable, institutions should implement strong encryption and anonymization techniques to mitigate risks of unauthorized access.
4. **Ensure Proper Consent and Notification:**

- Educational institutions must obtain clear, informed consent from parents or guardians before processing and transferring student data, particularly when it involves non-EU data centers or advertising services.

5. **Explore EU-based Cloud Alternatives:**
    - Where feasible, institutions should consider using EU-based CDN and cloud services that comply with GDPR and ensure data localization to avoid cross-border transfers.