An Empirical Study of Directory Service Dependencies

BENJAMIN OTHMER, University of Twente, The Netherlands

In this paper, an empirical analysis of LDAP services and their security-relevant properties as well as their concentration around the most popular service providers was conducted, by investigating scans of the Internet contained in snapshots of the months of 2022 in the Censys Universal Internet Dataset (CUID). "Off-the-shelf" data sets like the CUID provide researchers with new and prospective data, allowing further avenues of research into this topic. After processing, extracting, and categorizing the CUID data pertaining to LDAP services, we observed a high amount of weak TLS implementations and an overall bad security posture. Additionally, we observed that outdated and weak TLS versions and ciphers are being updated and replaced, but at a slow rate. We found a concentration of services around multiple large service providers in the United States, while LDAP services deployed in Europe concentrate on a few large ones. Services concentrate on service providers offering cloud- and customer-oriented solutions. A high number of possibly outsourced services and an indication of worse TLS deployment practices at providers with a high emphasis on customer-dependant implementations, such as cloud-oriented service implementations, can be seen. Limitations of the current CUID dataset, like a lack of TLS data concerning services deployed on port 389 were identified. Finally, data such as the CUID offers new avenues of research, and further properties of LDAP services over time could be investigated in (more extensive) future work.

CCS Concepts: • **Networks** → *Web protocol security*.

Additional Key Words and Phrases: LDAP, Active Directory, Internet Security, Dependencies

1 INTRODUCTION

The Lightweight Directory Access Protocol (LDAP), is a well established protocol deployed in information technology (IT) infrastructure. Since its inception, it has continuously evolved and adapted, recognized as a standard protocol enabling organizations to communicate between their internal and external directory services landscape [25]. This function in the organization's network and the required exposure to the public-facing Internet highlights the critical role of cybersecurity, especially concerning fulfilling the requirements of confidentiality, integrity, and availability of the service and network. In this paper, we aim to establish an overview of LDAP Services on the public Internet that have been outsourced to third-party service providers and what implications this has regarding the quality of the security of these services. This will be accomplished by analyzing Internet-wide scans, evaluating TLS implementations, names found in TLS certificates and DNS data, concentrations of services around service providers, and attempting to identify the number of outsourced services.

Modern service infrastructures increasingly rely on outsourcing their services to third parties such as Identity Providers (IdPs), Content Delivery Infrastructures (CDIs), and Content Delivery Networks (CDNs) [24]. Recent incidents, such as the Mirai-Dyn Distributed Denial of Service (DDoS) attack substantiate the magnitude of an overt reliance on dependencies. Third-party DNS CDNs of web services were attacked by a large botnet leading to cascading failures for any dependent service. Furthermore, transitive dependencies (dependencies with external dependencies) exacerbate the issue [21]. Malicious tools become increasingly available and accessible. The attack surface and the number of dependencies of web services are high [21], and malicious tools (such as botnets) become more available and accessible, increasing incidents [35].

Not only denials of service are a concern. LDAP services are vulnerable to injection attacks due to their database and searching capability nature: The Open Web Application Security Project (OWASP) ranks injection attacks among the top 3 most critical risks [8]. Additional vulnerabilities in the service itself can exist, as demonstrated by N. Syynimaa (2018), showcasing how an organization's Azure Active Directory (AAD) service and thus the dependent infrastructure can be accessed by leveraging vulnerabilities [33].

These vulnerabilities can have a severe impact, as can be seen when examining the SolarWinds hack (2020), coined the "largest and most sophisticated attack the world has ever seen" by Microsoft corporation president Brad Smith. This large-scale supply chain attack impacted up to 18,000 dependent organizations [2]. The authentication service was breached by exploiting a vulnerability in which cloud services accepted IdP access tokens and on-site tokens provided by AAD [17]. This is just one example of the compromise of an LDAP provider upon which many customers depend. A single point of failure implies that all these customers are at risk, independent of whether the LDAP service implementation itself is hacked or just the provider hosting it.

Despite these reoccurring incidents, no investigations into the security posture of LDAP services on the Internet were conducted thus far. The difficulty herein lies in orchestrating the collection and evaluation of representative data sets. In 2015, Censys published the Universal Internet Dataset (CUID), a comprehensive collection of scan data, providing snapshots of the public Internet, including information about devices, services, and protocols [15]. By analyzing snapshots of this data for 2022, we attempt to answer the following research questions:

1: What was the state of LDAP services TLS properties within the CUID, including TLS version, TLS ciphers, and certificate-signing practices in 2022?
2: How were LDAP services distributed across countries, organizations, and top-level domains (TLDs) within the CUID in 2022?
3: Who are the top service providers of LDAP services, and what are their characteristics regarding protocol security and service type, and can we establish any evidence for outsourcing?

Our main contributions are: (1) We establish a baseline for the use of LDAP services and give numbers for the deployment of LDAP services on the Internet. (2) We provide an analysis of the properties of the respective TLS deployments, including connection properties

TScIT 39, July 7, 2023, Enschede, The Netherlands

and certificates. (3) We map LDAP services to countries and organizations. (4) We investigate the top organizations hosting LDAP services, including TLS deployments and investigate outsourcing.

The following sections introduce the background of LDAP services and related work, document the methodology of analyzing the CUID, provide an overview of the data set and data categorization, and show the results of our findings. After discussing these findings, the conclusion and future work sections highlight further areas of research that were identified during the process.

2 BACKGROUND AND RELATED WORK

2.1 LDAP

LDAP is a client-server communication protocol, used to facilitate the transfer and access of data stored in directory services, such as Active Directory (AD). LDAP is standardized, as specified by the Internet Engineering Task Force (IETF) in their 2006 publication of RFC 4511 [31]. LDAPv3 remains open and compatible across directory services of multiple vendors and their respective versions such as AD, Azure AD (AAD), or Apache Directory Server.

2.2 Directory Services

Directory services are specialized databases, typically organized in a hierarchical, tree-like structure, and include the following use cases: (1) Authentication: Directory services are used to authenticate users and applications between the organization's network services with user- and service accounts. In addition to storing organizational information such on users and assets, access policies can be configured for users, applications, and groups [13]. (2) Organization Management: Due to the tree-like structure [30], LDAP is commonly used to hierarchically structure users in an organization in the database. Like an address system, it facilitates storing data in a tree structure, i.e. a corporation with its division, locations, teams, and room numbers [19]. Additionally, it enables other applications to interact with this data, such as e-mail services and computers. (3) Network Management: Storing i.e. user-, network resource, and device (such as printers) data. This allows for the management of devices, by mapping physical locations to network assets in combination with access management and authentication [19]. (4) Single Sign-On (SSO): An authentication method, that enabling access to multiple services and devices with the benefit of using single credentials. This prevents users from having to authenticate to every service under the same domain [32]. In general, multiple stakeholders such as organizations and corporations use directory (LDAP) services.

2.3 DNS

LDAP services are deployed on the public Internet by organizations, allowing for offsite access to internal resources by members or applications, but also third-party identity providers (IdP) including Content Delivery Infrastructure (CDI) [12]. IdPs and CDIs are actively used by organizations to outsource their infrastructure, including their LDAP services. DNS maps addresses to hostnames, and by querying the Fully Qualified Domain Name (FQDN) of the server, its address will be returned. This can also be done in reverse: Querying the Internet Protocol (IP) address returns the FQDN [34].

The FQDN is comprised of the subdomain (SD), second-level domain (SLD), and top-level domain (TLD), i.e.: `"auth.org.com"` is comprised of `"SD.SLD.TLD"`. Additionally, it enables organizations a separation of concerns about their services by assigning components of their domain to servers that fulfill that role. This facilitates multiple services sharing the same domain [34].

2.4 TLS

LDAP(S) supports TLS/SSL as an extension to fulfill information security requirements. The most recent edition of Transport Layer Security (TLS) is a standard protocol defined by the IETF, with the most recent version: TLS 1.3 released in the year 2018. Previous versions of TLS, excluding 1.2 and 1.3 have been declared deprecated [27] with TLS 1.2 being limited to a set of Ciphers suites such as Advanced Encryption Standard (AES) in Galois/Counter Mode (GCM) [29].

TLS, a cryptographic protocol, employs public key X.509 certificates, mapping an individual or organization to a public key by digital signature [11]. These certificates are either self-signed or signed by a certificate authority (CA), increasing their trustworthiness. The public key is then used in the TLS handshake between the client and server, to prove the authenticity of the server. The cipher suites that are supported for usage by TLS have varying levels of security. Insecure and outdated suites (specific to TLS) are classified as deprecated as per RFC 5246 [28].

2.5 Related Work

At this point, we are not aware of any research addressing the questions formulated in this paper. Specifically, LDAP services are not correlated with large data sets of Internet-wide scans.

Concerning SSL/TLS security on the Internet, in 2011 Holz et al. conducted active and passive measurements of X.509 certificates used in SSL/TLS authentication purposes, identifying multiple issues with the X.509 Public Key Infrastructure (PKI) such as incorrect identification chains of CA's and highlights the prevalence of self-signed certificates in use [18].

In the year 2020, A. Kashaf et al. investigated the prevalence and impact of DNS, CDN, and certificate revocation checking third-party CAs dependencies. They observed an increase of 1% to 5% in critical dependencies on websites and an increase in the concentration of these dependencies around service providers [21].

Doan et al. observed that the number of dependencies on CDIs nearly doubled for .com, .net, and .org domains between the years 2015 and 2022 [14]. The consequences of a high dependency of domains on third parties have been demonstrated in the war in Ukraine. Jonker et al. observed a high reliance on Western CAs, which must now rapidly be replaced by Russian ones [20].

Concerning TLS deployment on a high level, Kotzias at al. analyzed passive measurements of the Internet between the years 2012 to 2018. They observed a general increase in replacing outdated TLS Versions (1.0 and 1.1) and the appearance of 1.3 (released in 2018). Additionally, they noted a decrease in connections utilizing RSA, being replaced by ECDHE for TLS key exchange [22].

3 METHODOLOGY

3.1 Data

To observe and measure these publicly accessible LDAP services, Internet-wide scans must be conducted. Typically, empirical researchers implement data-collection processes for research purposes. This involves facing technical, legal, and ethical considerations to answer even simple questions. Censys, a project that regularly conducts and provides horizontal scans of applications in the public IPv4 address space provides an off-the-shelf measurement of the Internet for researchers.

Censys scans the public address space regularly using a combination of tools. To identify active hosts, ZMap scans IP addresses in network ranges, probing addresses, and ports of services that are responsive [15]. These hosts are consecutively forwarded in the data processing pipeline to an application scanner: ZGrab, which will attempt to establish a TCP connection with the addresses and services, optionally initiating StartTLS (dependent on the service). Once the handshake is initiated, ZGrab parses and extracts relevant information from the TLS negotiation process, such as certificates obtained, including (but not limited to) versioning, cipher suites, certificate's issuer, subject, and validity [16]. This data is then indexed and stored.

Data sets resulting from these Censys scans are published frequently. For this research, eleven Universal Internet scan data sets, conducted between February and December 2022 were combined and evaluated. The sets were selected from the middle of the month, depending on the availability of the data, starting in February. This is when Censys commenced including scans of LDAP applications not only limited to port 389. LDAP can also be deployed on other ports, outside of the officially designated ones specified by the Service Name and Transport Protocol Port Number Registry [5].

Since we aim to investigate the distribution of LDAP Services by organizations and countries, on the internet - further data is required: (1) To map IP addresses to ASNs, Border Gateway Protocol (BGP) data provided by the Route Views Project, for the same time span as the CUID snapshots, were used [4]. (2) To map ASNs to organizations that participate in internet routing, data sets originating from the Center for Applied Internet Data Analysis (CAIDA) for the same period as the Censys data set were used [1].

3.2 Tools

To download and process the very large Censys data set, usage of the University's research computer cluster was essential. After downloading the Internet scan data sets in Avro (row-oriented) format and converting them to Parquet (column-oriented) format to achieve satisfactory data processing time, they were retrieved and merged using PySpark, resulting in one combined data set for every sample. PySpark, the Python API for Apache Spark is a distributed computing framework suitable for such tasks. After basic filtering and transforming of the data were completed, we used Pandas to analyze the data set. The PyAsn library, designed for analyzing and processing ASN information is deployed to analyze the IP addresses and used to map the IP addresses to ASNs and ASNs to organizations [7]. Incorporating these two data sets combined with the CUID enables us to gain insights into the organizational ownership and network infrastructure distribution of LDAP services, complementing the TLS property analysis. The tldextract Python module is used to parse domain names, especially for extracting SLDs and TLDs, after initial attempts using regular expressions provided inconsistent results [23].

3.3 Transformation and Extraction

First, to reduce the overall size of the data set, the relevant columns are selected. These columns include the IPv4 host identifier, a list of DNS names, a list of reverse-DNS names, a list of service names running on the host, the ports of the services running on the host, TLS data: Versions, ciphers, and certificate leaf data, including columns documenting whether the certificate is self-signed, the signature algorithm and names.

After filtering and reformatting the data set into a Pandas data frame, multiple operations on the data set are performed, to enrich the data set: (1) Extracting the domain components of the DNS, reverse-DNS records, and certificate names using the tldextract library. The URLs are parsed, such that erroneous (i.e., `HOSTNAME`), and local addresses are excluded. (2) The IP addresses of the hosts are mapped to ASNs, these ASNs are then mapped to organization IDs, which are then mapped to organization names and locations (countries) by using the Routeviews BGP data extracted and accessed via PyASN and CAIDA data sets. (3) The TLS ciphers are extracted into multiple columns, to provide further insight into the key exchange, authentication, encryption, and hashing mechanisms. A column containing a boolean value about the classification of "weak cipher" is computed. More details on the classification mechanisms will be explained in the following section. Finally, in this phase, the date of the Internet scan is added to the data set which is then stored as a CSV file.

4 DATA SET AND DATA CATEGORIZATION

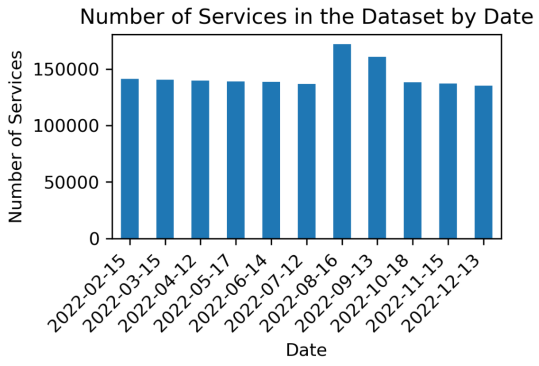

Fig. 1. Number of LDAP Services by Date in 2022

This section will elaborate on the process of evaluating the data to answer each of the research questions. After the data set is initially transformed and extracted (as described in the previous section), the resulting CSV files are loaded and concatenated using Pandas.

The columns about the service properties of each host contain arrays that must first be retrieved and restructured. This is achieved

by exploding the full data set, and applying a final filter on the service name, resulting in a Pandas data frame. As can be seen in Fig. 1, after filtering the hosts identified as hosting LDAP services, with a mean of M = 143,539 LDAP services (SD: 11,752) can be found per sample [3].

To investigate the TLS properties (Version, Cipher suites, and certificate signing practices), first of all, the distinct TLS versions found in the data set are classified by deprecation status, where all versions except 1.2 and 1.3 are classified as deprecated, as specified by RFC 5246 [27]. To accomplish this, the service must deploy TLS (or it must have been detected). Further investigations into the distribution of ports provide some preliminary insights into the limitations of the CUID.

Table 1. Detection of Port Usage for LDAP Service Hosts with TLS data

Port	Count	Port	Count	Port	Count
636	462985	6360	11	800	2
6636	11	3636	8	641	2
443	11	801	7	1636	1
700	11	30636	5	6366	1

As can be seen in Table 1, none of the LDAP services deployed on port 389 (the traditional "plaintext" port) contain any data on TLS features. When investigating the TLS properties, only columns where TLS is detected can be evaluated, in which the LDAP service port 389, of which 70.7% of all services run, will not be included. This aligns with the TLS detection statistics: Overall, the combined data set includes 463,055 (29.3%) TLS deployments.

Table 2. "Weak" TLS Cipher Suites

Cipher
RSA_WITH_AES_128_CBC_SHA
RSA_WITH_AES_256_CBC_SHA
RSA_WITH_AES_128_GCM_SHA256
RSA_WITH_AES_256_GCM_SHA384
RSA_WITH_3DES_EDE_CBC_SHA
ECDHE_RSA_WITH_AES_128_CBC_SHA
ECDHE_RSA_WITH_AES_256_CBC_SHA

The distinct ciphers found in use by the hosts are grouped and evaluated individually. The distinct "weak" ciphers found in the data sets with their corresponding versions can be seen in Table 2. The classification of these cipher suites depends on the relevant RFCs that define and/or deprecate cipher suites.

In general, cipher suites utilizing the Rivest Shamir Adleman algorithm (RSA) for key exchange and authentication are not supported in TLS version 1.3. For version 1.2, they are classified as "weak" and will soon officially be deprecated, as can be seen in a "work in progress" draft by the Internet Engineering Task Force (IETF). The reason for this classification is the algorithm is non-ephemeral, and thus does not support Perfect Forward Secrecy (PFS) [10]. Cipher suites using Secure Hash Algorithms Version 1 (SHA-1) are classified as weak because SHA-1 is generally considered vulnerable and was deprecated and disallowed by the National Institute for Standards and Technology (its original author) in the year 2011 with SP 800-131A [9]. Cipher suites employing Cipher Block Chaining (CBC) mode [6] are also classified as weak (in TLS version 1.2) due to it having vulnerabilities to padding attacks and not providing integrity protection for the encrypted data. Galois/Counter Mode (GCM) fixes these specific vulnerabilities and should thus be used in place of CBC.

The corresponding field in the CUID to determine if the certificate is self-signed is computed by comparing issuer- and subject names, and provided in the data set. If both match, Censys automatically categorizes the certificate as self-signed [11].

To evaluate the distribution of services the IP addresses were mapped to organizations and countries. Each match uses CAIDA and Routeviews data sets nearest to the date of the scan to obtain results as accurately as possible. It was possible to match 99.97% of the addresses from the data set except 439 due to an underlying issue with the implementation of the interface that was developed for this analysis.

After determining the top service providers that LDAP services are concentrated around, these providers will be analyzed in more depth: We will attempt to classify whether they act as third parties (outsourcing destinations) for LDAP Services and whether they are customer-oriented or service-oriented. Accurately determining whether a service is outsourced or not, is hard without further insights into the properties of the service, which lies outside the scope of the data set. We have devised two methods (M. 1 and M. 2) to determine evidence indicating whether a service *could* be outsourced:

M. 1: Let X := {Leaf Certificate Names}
Y := {DNS Names} $\cup$ {Reverse-DNS Names}
$\forall x \in X$: If $x \in Y$, evidence is found.
If $x \notin Y$, no evidence is found.

M. 2: Let X := {Reverse-DNS Names}
Y := {DNS Names}
$\forall x \in X$, if $x \notin Y$, evidence is found.
If $x \in Y \wedge |Y \setminus X| > 0$, evidence is found.
If $x \in Y \wedge |Y \setminus X| = 0$, no evidence is found.

In both (1) and (2), X and Y, each element is stripped into `"SLD.TLD"` (as explained in 2.3). Both methods (individually but also in combination) can provide insight into whether an LDAP service might be outsourced to a third party (outside of the original domains), but can respectively only be applied to a limited portion of the data set. Table 3 provides an overview of how many samples in the total data set each method could be applied to respectively. For methods 1 and 2, if either X or Y are "None", the item is classified as "invalid" (Not Applicable):

Table 3. Application of Methods to Gather Evidence for Outsourcing

Category	Method 1	Method 2
Applicable	242,831 (15.4%)	813,052 (51.5%)
Not Applicable	1,336,107 (84.6%)	765,886 (48.5%)

To determine if a service is possibly outsourced, the union of methods 1 and 2 is calculated on the subset of the intersection

of applicable rows for both methods. With the available data, the combined methods can be applied to 183,687 (11.6%) of the samples.

Additionally, the characteristics of the service provider organizations will be investigated, by looking at the structure of the FQDNs, i.e., if the naming of the subdomains indicates whether the host is cloud-service oriented or customer-oriented. The latter hosts "ready-to-go" servers (solutions) that come pre-configured to a certain extent and traditionally deploy multiple services, requiring less configuration by the customer. Cloud-service providers on the other hand are typically more "off-hand" - providing the infrastructure to be configured by the customer.

5 RESULTS

5.1 The State of TLS for LDAP Services

A mean of $M = 143,539$ ($SD = 11,752$) LDAP Services were found in the scans each month. The first research question investigates the following TLS properties, which are summarized and displayed in table 4 (*Note*: Weak or Deprecated TLS extensions are highlighted in red, otherwise in green.):

Table 4. TLS Properties of LDAP Services in 2022

Category	Feb-Apr	May-Aug	Sep-Dec
Version			
TLSv1_0	7,263 (05.7%)	8,187 (04.9%)	7,066 (04.2%)
TLSv1_1	839 (00.7%)	993 (00.6%)	891 (00.5%)
TLSv1_2	93,265 (72.7%)	120,216 (71.5%)	116,101 (69.7%)
TLSv1_3	27,000 (21.0%)	38,692 (23.0%)	42,542 (25.5%)
Encryption			
3DES-EDE-CBC	1,104 (00.9%)	1,178 (00.7%)	975 (00.6%)
AES128-CBC	9,416 (07.3%)	11,080 (06.6%)	9,031 (05.4%)
AES256-CBC	19,183 (14.9%)	23,676 (14.1%)	21,608 (13.0%)
AES128-GCM	34,134 (26.6%)	44,811 (26.7%)	44,308 (26.6%)
AES256-GCM	45,192 (35.2%)	61,073 (36.3%)	64,007 (38.4%)
CHACHA02-P1305	19,338 (15.1%)	26,270 (15.6%)	26,671 (16.0%)
Key Exchange			
RSA	29,241 (28.8%)	36,854 (28.5%)	35,428 (28.6%)
ECDHE	72,126 (71.2%)	92,542 (71.5%)	88,630 (71.4%)
Sign. Algorithm			
ECDSA-SHA1	3 (00.0%)	4 (00.0%)	4 (00.0%)
MD5-RSA	875 (00.8%)	942 (00.6%)	808 (00.5%)
SHA1-RSA	20,177 (18.5%)	34,534 (20.6%)	33,629 (20.2%)
ECDSA-SHA256	187 (00.2%)	295 (00.2%)	322 (00.2%)
ECDSA-SHA384	480 (00.4%)	780 (00.5%)	722 (00.4%)
ECDSA-SHA512	17 (00.0%)	25 (00.0%)	24 (00.0%)
SHA256-RSA	84,895 (78.0%)	128,039 (76.2%)	127,476 (76.5%)
SHA256-RSAPSS	17 (00.0%)	42 (00.0%)	51 (00.0%)
SHA384-RSA	1,340 (01.2%)	1,871 (01.1%)	1,972 (01.2%)
SHA512-RSA	870 (00.8%)	1,545 (00.9%)	1,583 (01.0%)
Signature			
Self-Signed	22,903 (21.0%)	37,485 (22.3%)	38,108 (22.9%)
Not Self Signed	85,965 (79.0%)	130,603 (77.7%)	128,492 (77.1%)

Version: The most common TLS version in use is 1.2 with a share of 72.66%. This proportion of the share declines over the year, ending at 69.68%. TLS version 1.3 starts with the second-highest share of 21.03% and ends with 25.53%, being the only version with a net growth of 21.4%. TLS version 1.0 is the third largest shareholder of being deployed on LDAP services at 5.66%, reducing in size over the year to 4.24%. Version 1.1 has the smallest share of services, declining across the year from 0.65% to 0.53%. **Key Takeaways:** Predominantly TLS versions 1.2 and 1.3 are deployed, 1.3 being the only version with a net growth.

Encryption: The encryption used by the TLS implementations in the data set was primarily (Advanced Encryption Standard) AES-based using GCM. With 35.21%, AES 256-bit (GCM) has the largest share, followed by the 128-bit implementation at 26.59%. The 256-bit variation grows to 38.42%, while the 128-bit variant remains constant. ChaCha20-Poly1305 is used in 15.06% and grows to 16.01%. The remaining versions (overall 23.14%) all use CBC. While AES 256-bit has the largest share in this category, CBC is problematic in TLS version 1.2 as discussed in section 4. Overall, the total share of this category declines to 18.98% by the end of the year. **Key Takeaways:** In general, strong encryption is deployed. Weak versions (i.e., using CBC) are gradually being phased out.

Key Exchange, Signature algorithm, and self-signed: For key exchange, ECDHE is majorly used, with a proportion of 71.2% compared to RSA at 28.8%. Over the year, ECDHE grows to 71.4% and RSA reduces to 28.6%. Examining the shares of signature algorithms, we notice that The proportion of self-signed certificates is 21.04%, increasing to 22.88% over the year. **Key Takeaways:** Overwhelmingly, ECDHE is used for Key Exchange. One out of five certificates is self-signed.

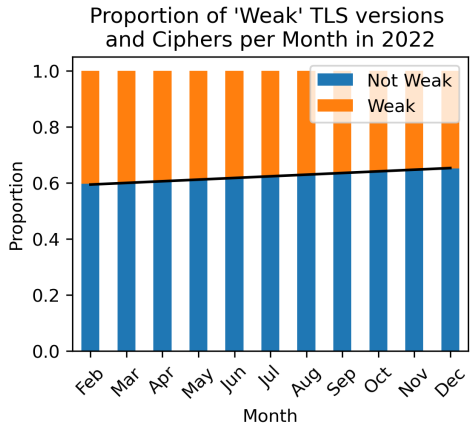

Fig. 2. Proportion of 'Weak' TLS implementations per Month

Fig. 2 shows the proportion of weak TLS versions deployed for LDAP Services. Starting with a share of 40.25%, the "weak" category slowly declines to 34.82% by the end of the year. The trendline indicates that while these "weak" implementations are being replaced, at this rate it will take several years. as is additionally highlighted by the trendline. **Key Takeaways:** Weak TLS implementations are being replaced at a slow rate.

5.2 The Distribution of LDAP Services

The left graph titled "Total Number of LDAP Services per Country" (included in Fig. 3), shows that the biggest proportion of LDAP

Services hosts is located in the United States of America (US) at 26.34%. Examining the top 10 providers, Germany (DE) with 10.77% lies in second place, followed by France (FR) at 5.56%, Poland (PL) at 5.29%, China (CN) at 3.78%, Russia (RU) at 3.66%, Brazil (BR) at 2.95%, the United Kingdom (GB) at 2.82%, Taiwan (TW) at 2.5%, and Italy (IT) at 2.27%. The distribution is very uneven, with a long tail. In total, the data set contained a total of 210 country code top-level domains (ccTLDs).

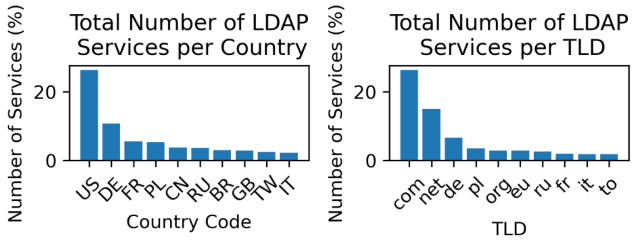

Fig. 3. Top 10 Number of LDAP Services per Country and TLD

Fig. 3 also shows the distribution of the Total Number of LDAP Services per TLD. This includes TLDs of LDAP services and DNS names. The graph shows the commercial TLD: 'com' having the largest share of 26.33%, followed by 'net', intended to only hold hosts of network providers at 14.91%. Next, the German 'de' at 6.57%, followed by the Polish "pl" at 3.56%, the general 'org' (for organizations) at 2.85%, European Union with 'eu' at 2.8%, Russia with 'ru' at 2.6%, France with 'fr' at 1.92%, Italy with 'it' and Tonga with 'to' at 1.8% [26]. **Key Takeaways:** The distribution of TLDs and countries is highly skewed towards US providers (26.34%) with the largest proportion by far, followed by Germany (10.77%) and France (5.56%).

Next, Fig. 4 shows the distribution of services among the top 8 organizations. Overall, several 12,614 distinct organizations were found in the data set, and while the percentages are smaller, relative to the country and TLD graphs, Amazon.com Inc. by itself runs 73,297 services on their servers, at 4.64% of the share.

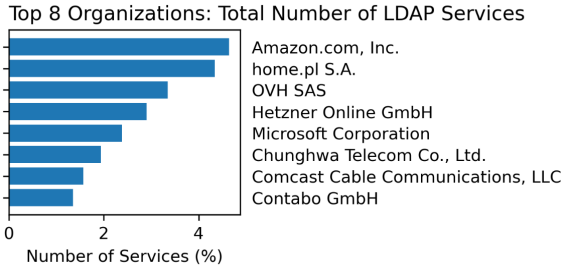

Fig. 4. Top 8 Organizations (Service Providers), Number of LDAP Services

5.3 The Top LDAP Service Providers (Organizations)

The top 4 service providers are derived from the concentration of services by organizations as depicted in Fig. 4. The list includes Amazon.com, Inc., home.pl S.A., OVH SAS, and Hetzner Online GmbH. For these providers, all services are hosted on either Port 389 or 636. In each case, we have identified between 45,831 to 73,297 samples. While this may still be a biased sample, the size is relatively high and should be representative.

Table 5. LDAP Services TLS Properties in 2022

Category	Amazon	home.pl	OVH SAS	Hetzner
Country	U.S.	Poland	France	Germany
Port: 389	70%	100%	76%	65%
Port: 636	30%	0%	24%	35%

The most popular reverse-DNS names Amazon hosts dominantly include "compute.amazonaws.com". Amazonaws is a cloud computing services-oriented provider, focussing on micro-service architectures and is categorized as "cloud-oriented". For home.pl S.A., "cloudserver" is dominantly included in the reverse DNS name. Overall, home.pl is a more customer-oriented solution, primarily offering ready-to-go server solutions, and thus is categorized as such. OVH SAS and Hetzner Online GmbH are categorized as customer-oriented because they mainly employ customer-oriented server solutions in their service portfolio.

Table 6. LDAP Services TLS Properties in 2022

Method	Amazon	OVH SAS	Hetzner Online
Self-signed*	2239 (10.5%)	2118 (16.9%)	1994 (12.9%)
TLS "Weak"	3391 (16.1%)	2436 (18.0%)	1757 (10.8%)
Method 1	10090 (59.0%)	1977 (20.9%)	2013 (15.3%)
Method 2	18853 (33.7%)	19608 (43.4%)	21049 (49.7%)
Combined			
True	15504 (97.5%)	5651 (65.8%)	7464 (60.4%)
+ TLS "Weak"	2166 (14.0%)	880 (15.6%)	647 (08.7%)
False	393 (2.5%)	2944 (34.3%)	4901 (39.6%)
+ TLS "Weak"	98 (24.9%)	433 (14.7%)	503 (10.3%)

Since home.pl hosts are exclusively discovered on port 389, the CUID does not include any information pertaining to TLS properties of the services. The comparison of the TLS properties between the three companies is shown in Table 6. LDAP Services deployed by Amazon have the lowest share of self-signed TLS certificates (10.49%). Combining both methods 1 and 2 to determine if Amazon is used for outsourced LDAP services gives Amazon the highest share of outsourced services at 97.53%. The sample size for "not-outsourced" LDAP Services is very low at only 98. When examining OVH SAS, it has the highest proportions of "weak" TLS Ciphers and Versions at 17.98% and self-signed TLS certificates at 16.93%. Hetzner, the provider with the lowest proportion of outsourced services according to the combined methods, with a total of 60.36% of services being outsourced has the lowest proportion of "weak" TLS certificates and versions, regardless of outsourcing. Overall, when comparing the proportion of weak certificates with respect to outsourcing, no significant changes in TLS "weakness" can be observed. **Key Takeaways:** Amazon has the highest number of detected outsourcing, fewer self-signed certificates but ca. 50% more "weak" TLS deployments than the other providers.

When inspecting the issuer's Common Name (CN) of the TLS certificates for each company, we observed that in general, Let's Encrypt had the largest share with 10,711 (25%), followed by DigiCert at 3,237 (7.5%) and Sectigo at 3,040 (7.4%). Services hosted by Amazon's infrastructure uses primarily Let's Encrypt (23.4%), and Amazon CA (15.4%). Services hosted by Hetzner Online GmbH overwhelmingly use Let's Encrypt with a share of 4,946 (46.6%) and Sectigo at 1114 (10.5%). Similarly, OVH SAS hosted services mainly use Let's Encrypt at 2066 with a share of (29.2%), Sectigo at 908 (12.2%), and GoDaddy at 213 (3%). **Key Takeaways:** Let's Encrypt is extensively used to sign certificates, services hosted by Amazon often use Amazon CA.

6 DISCUSSION

LDAP Services TLS Properties in 2022 were concerning. Most (93.69%) of deployed TLS Versions were not deprecated, and version 1.3 deployment is growing, as predicted by Kotzias at al. [22], and overall strong encryption algorithms were used. When observing the combination of overall TLS version and cipher suite "weakness", a staggering amount of 40.25% of services fulfilled these criteria. While this proportion reduced to 34.82% by the end of the year, it will take several more years to replace them at this rate. High numbers of deployments of CBC and RSA can be observed. Considering that LDAP services are primarily used for authentication purposes, many services are susceptible to attacks on the version and cipher suites alone.

Overall, there is a highly skewed distribution of LDAP Services with a growing concentration around service providers in the United States. Germany, France, and Poland are the main service providers in Europe. Examining the distribution of TLDs among services show that a lot of addresses resolve to the commercial and network provider sectors. The distribution of TLDs is similar to what Doan et al. observed: a high number of CDIs host under the `.com`, `.net`, and `.org` domains [14]. From the distribution of services per organization, it can be seen that Microsoft, Amazon, and Comcast are among the top 10 service providers, but with much lower market shares in proportion to LDAP services per country of the respective organization. This implies that a variety of U.S. based providers (Amazon having the highest share), are extensively used for hosting LDAP services.

When examining the top service providers in more detail, some differences become immediately apparent. The distribution of ports across providers appears to be similar, except home.pl S.A., where the data set counts services running exclusively on port 389. Observing the results, a difference between service providers, including their deployment model of services and hosts can be seen: Amazon (hosting the majority of LDAP Services found in the CUID), is primarily cloud-hosting oriented, with a microservice architecture. Nearly all (97.53%) of LDAP Services appear to be "outsourced". These services also have the highest number of certificates signed by CAs, with the Amazon CA having a large share of the CA dependencies. OVH SAS also appears to have a large portion of "outsourced" services, itself also advertising cloud-based microservices. This is in stark contrast to Hetzner Online GmbH having primarily a customer-oriented model, providing server solutions that host multiple services simultaneously. Amazon and OVH SAS have the highest number of "weak" TLS implementations with an increase of over 50% in comparison to Hetzner Online GmbH. These outsourced services could potentially be a target of attacks due to weak TLS configuration [6]. This could also be an indication that providers like Hetznerm offering "ready-to-go" virtual server solutions, are compliant than custom-configured cloud services. Finally, all services hosted by these providers rely on CAs such as Let's Encrypt, Digicert and Setigo.

6.1 Limitations

The combination of these data sets and the evaluations face multiple limitations. The data is derived primarily from the CUID. The scans were obtained each month of the year 2022. When examining the distribution of ports for LDAP services, Censys only scanned for LDAP on port 389 in January. After January, LDAP was also detected on other ports, but scanned differently than the services detected running on port 389: None of these service scans include data about their TLS properties. This bias in scanning techniques significantly limits the results, considering that this port is the most popular port for LDAP service deployments. For ethical reasons, organizations have the option to "opt out" of the scans, excluding them from the data set. Outside of the published code from Censys, it is also difficult to determine how to correct their methods function in detail, which is another factor of which the impact is unknown. Additional to the CUID, we also used BGP data from the Routeviews [4] and CAIDA projects [1]. The limitation herein mainly lies in the correlation of the Censys data set with the other data sets. IP addresses are not static and can be re-assigned regularly. The IP addresses from the CUID snapshots were correlated with IP addresses from the other data sets, closest matching to the Censys data set by date. This introduces uncertainty in the time since a portion of the IP Addresses could have been assigned to different organizations in the delta of days between the Censys scan and CAIDA and Routeviews data sets that we were able to access. Finally, we were not able to evaluate the validity of the certificates itself, which in future work should be considered.

7 CONCLUSION AND FUTURE WORK

We have investigated LDAP service dependencies in CUID snapshots of the Internet in 2022. We were able to observe a generally weak posture of TLS implementations (using deprecated TLS Versions or weak cipher suites). While those weak implementations are being upgraded and replaced, it will require several more years until completion at this rate. A high number of services concentrate around multiple service providers, which could lead to significant impacts on the availability of LDAP services, were one of those service providers to be affected by an outage or breach of security. After investigating the most popular service providers hosting LDAP services, we observed that all providers included services with "weak" TLS implementations. The differences in the top providers were mainly observed when categorizing them by business and deployment model. Amazon is majority focused on "cloud-hosted" deployments, while Hetzner Online GmbH is more focused on "customer-oriented" deployments. This data might also indicate

that the two more "cloud" and "outsourcing" oriented providers face higher issues with weak TLS implementations, in comparison to their counterparts with a higher share of "ready-to-go" and less "cloud-hosted" service deployments.

Finally, this type of research can be expanded significantly in multiple aspects: it would be beneficial to investigate these scans over a longer period, and use different data sets to give a more enhanced perspective. IP's identified during this process could manually be scanned in-house to circumvent the previously mentioned limitations. Furthermore, the CUID scans themselves could be improved by including TLS data about services running on port 389, which would provide a significant upgrade in gaining a "high-level" insight into the most popular port for LDAP service deployment. Overall, TLS properties themselves could be investigated in more depth, i.e., by measuring properties of TLS certificates such as validity, key lengths, and signature chains. Particularly, determining if services are outsourced or not requires a more general and precise approach.

REFERENCES

[1] 2014. Inferred AS to Organization Mapping Dataset. https://www.caida.org/catalog/datasets/as-organizations/

[2] 2021. SolarWinds hack was 'largest and most sophisticated attack' ever: Microsoft president. *Reuters* (Feb. 2021). https://www.reuters.com/article/us-cyber-solarwinds-microsoft-idUSKBN2AF03R

[3] 2023. pandas documentation — pandas 2.0.2 documentation. https://pandas.pydata.org/docs/

[4] 2023. Route Views – University of Oregon Route Views Project. https://www.routeviews.org/routeviews/

[5] 2023. Service Name and Transport Protocol Port Number Registry. https://www.iana.org/assignments/service-names-port-numbers/service-names-port-numbers.xhtml?search=ldap

[6] N. J. Al Fardan and K. G. Paterson. 2013. Lucky Thirteen: Breaking the TLS and DTLS Record Protocols. In *2013 IEEE Symposium on Security and Privacy*. IEEE, Berkeley, CA, 526–540. https://doi.org/10.1109/SP.2013.42

[7] Hadi Asghari. 2023. pyasn. https://github.com/hadiasghari/pyasn original-date: 2014-05-22T10:11:09Z.

[8] Matthew Bach-Nutman. 2020. Understanding The Top 10 OWASP Vulnerabilities. https://doi.org/10.48550/arXiv.2012.09960 arXiv:2012.09960 [cs].

[9] Elaine Barker and Allen Roginsky. 2019. *Transitioning the use of cryptographic algorithms and key lengths.* Technical Report NIST SP 800-131Ar2. National Institute of Standards and Technology, Gaithersburg, MD. NIST SP 800–131Ar2 pages. https://doi.org/10.6028/NIST.SP.800-131Ar2

[10] Carrick Bartle and Nimrod Aviram. 2023. *Deprecating Obsolete Key Exchange Methods in TLS 1.2.* Internet Draft draft-ietf-tls-deprecate-obsolete-kex-02. Internet Engineering Task Force. https://datatracker.ietf.org/doc/draft-ietf-tls-deprecate-obsolete-kex Num Pages: 20.

[11] Sharon Boeyen, Stefan Santesson, Tim Polk, Russ Housley, Stephen Farrell, and David Cooper. 2008. *Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile.* Request for Comments RFC 5280. Internet Engineering Task Force. https://doi.org/10.17487/RFC5280 Num Pages: 151.

[12] Deland-Han. 2023. Enable Lightweight Directory Access Protocol (LDAP) over Secure Sockets Layer (SSL) - Windows Server. https://learn.microsoft.com/en-us/troubleshoot/windows-server/identity/enable-ldap-over-ssl-3rd-certification-authority

[13] Brian Desmond, Joe Richards, Robbie Allen, and Alistair G. Lowe-Norris. 2008. *Active Directory: Designing, Deploying, and Running Active Directory.* "O'Reilly Media, Inc.".

[14] Trinh Viet Doan, Roland van Rijswijk-Deij, Oliver Hohlfeld, and Vaibhav Bajpai. 2022. An Empirical View on Consolidation of the Web. *ACM Transactions on Internet Technology* 22, 3 (Feb. 2022), 70:1–70:30. https://doi.org/10.1145/3503158

[15] Zakir Durumeric, David Adrian, Ariana Mirian, Michael Bailey, and J. Alex Halderman. 2015. A Search Engine Backed by Internet-Wide Scanning. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. ACM, Denver Colorado USA, 542–553. https://doi.org/10.1145/2810103.2813703

[16] Zakir Durumeric, Eric Wustrow, and J. Alex Halderman. 2013. {ZMap}: Fast Internet-wide Scanning and Its Security Applications. 605–620. https://www.usenix.org/conference/usenixsecurity13/technical-sessions/paper/durumeric

[17] Guido Grillenmeier. 2021. Now's the time to rethink Active Directory security. *Network Security* 2021, 7 (July 2021), 13–16. https://doi.org/10.1016/S1353-4858(21)00076-3

[18] Ralph Holz, Lothar Braun, Nils Kammenhuber, and Georg Carle. 2011. The SSL landscape: a thorough analysis of the x.509 PKI using active and passive measurements. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. ACM, Berlin Germany, 427–444. https://doi.org/10.1145/2068816.2068856

[19] Nalini C Iyer, Anil M. Kabbur, and Heera G. Wali. 2020. Implementation of Active Directory for efficient management of networks. *Procedia Computer Science* 172 (Jan. 2020), 112–114. https://doi.org/10.1016/j.procs.2020.05.016

[20] Mattijs Jonker, Gautam Akiwate, Antonia Affinito, Kc Claffy, Alessio Botta, Geoffrey M. Voelker, Roland Van Rijswijk-Deij, and Stefan Savage. 2022. Where .ru?: assessing the impact of conflict on russian domain infrastructure. In *Proceedings of the 22nd ACM Internet Measurement Conference*. ACM, Nice France, 159–165. https://doi.org/10.1145/3517745.3561423

[21] Aqsa Kashaf, Vyas Sekar, and Yuvraj Agarwal. 2020. Analyzing Third Party Service Dependencies in Modern Web Services: Have We Learned from the Mirai-Dyn Incident?. In *Proceedings of the ACM Internet Measurement Conference*. ACM, Virtual Event USA, 634–647. https://doi.org/10.1145/3419394.3423664

[22] Platon Kotzias, Abbas Razaghpanah, Johanna Amann, Kenneth G. Paterson, Narseo Vallina-Rodriguez, and Juan Caballero. 2018. Coming of Age: A Longitudinal Study of TLS Deployment. In *Proceedings of the Internet Measurement Conference 2018 (IMC '18)*. Association for Computing Machinery, New York, NY, USA, 415–428. https://doi.org/10.1145/3278532.3278568

[23] John Kurkowski. 2023. tldextract: Accurately separates a URL's subdomain, domain, and public suffix, using the Public Suffix List (PSL). By default, this includes the public ICANN TLDs and their exceptions. You can optionally support the Public Suffix List's private domains as well. https://github.com/john-kurkowski/tldextract

[24] Craig Labovitz, Scott Iekel-Johnson, Danny McPherson, Jon Oberheide, and Farnam Jahanian. 2010. Internet inter-domain traffic. In *Proceedings of the ACM SIGCOMM 2010 conference (SIGCOMM '10)*. Association for Computing Machinery, New York, NY, USA, 75–86. https://doi.org/10.1145/1851182.1851194

[25] Greg Lavender and Mark Wahl. 2004. Internet Directory Services Using the Lightweight Directory Access Protocol. In *The Practical Handbook of Internet Computing*, Munindar Singh (Ed.). Vol. 20042960. Chapman and Hall/CRC. https://doi.org/10.1201/9780203507223.pt4 Series Title: Chapman & Hall/CRC Computer & Information Science Series.

[26] Jon Postel. 1994. *Domain Name System Structure and Delegation.* Request for Comments RFC 1591. Internet Engineering Task Force. https://doi.org/10.17487/RFC1591 Num Pages: 7.

[27] Eric Rescorla. 2018. *The Transport Layer Security (TLS) Protocol Version 1.3.* Request for Comments RFC 8446. Internet Engineering Task Force. https://doi.org/10.17487/RFC8446 Num Pages: 160.

[28] Eric Rescorla and Tim Dierks. 2008. *The Transport Layer Security (TLS) Protocol Version 1.2.* Request for Comments RFC 5246. Internet Engineering Task Force. https://doi.org/10.17487/RFC5246 Num Pages: 104.

[29] Joseph A. Salowey, David McGrew, and Abhijit Choudhury. 2008. *AES Galois Counter Mode (GCM) Cipher Suites for TLS.* Request for Comments RFC 5288. Internet Engineering Task Force. https://doi.org/10.17487/RFC5288 Num Pages: 8.

[30] Martin Scheller, Klaus-Peter Boden, Andreas Geenen, and Joachim Kampermann. 1994. X.500. In *Internet Werkzeuge und Dienste: Von „Archie" bis „World Wide Web"*, Martin Scheller, Klaus-Peter Boden, Andreas Geenen, and Joachim Kampermann (Eds.). Springer, Berlin, Heidelberg, 129–140. https://doi.org/10.1007/978-3-642-85137-7_9

[31] Jim Sermersheim. 2006. *Lightweight Directory Access Protocol (LDAP): The Protocol.* Request for Comments RFC 4511. Internet Engineering Task Force. https://doi.org/10.17487/RFC4511 Num Pages: 68.

[32] D. Subbarao, Bhagya Raju, Farha Anjum, Ch venkateswara Rao, and B. Mahender Reddy. 2023. Microsoft Azure active directory for next level authentication to provide a seamless single sign-on experience. *Applied Nanoscience* 13, 2 (Feb. 2023), 1655–1664. https://doi.org/10.1007/s13204-021-02021-0

[33] Nestori Syynimaa. 2018. Who Would you Like to be Today?: Impersonation by Fake Azure Active Directory Identity Federation. In *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*. IEEE, New York, NY, USA, 1598–1604. https://doi.org/10.1109/TrustCom/BigDataSE.2018.00232

[34] Olivier van der Toorn, Moritz Müller, Sara Dickinson, Cristian Hesselman, Anna Sperotto, and Roland van Rijswijk-Deij. 2022. Addressing the challenges of modern DNS a comprehensive tutorial. *Computer Science Review* 45 (Aug. 2022), 100469. https://doi.org/10.1016/j.cosrev.2022.100469

[35] Polly Wainwright and Houssain Kettani. 2019. An Analysis of Botnet Models. In *Proceedings of the 2019 3rd International Conference on Compute and Data Analysis (ICCDA 2019)*. Association for Computing Machinery, New York, NY, USA, 116–121. https://doi.org/10.1145/3314545.3314562