# An infographic showing all URL parts, including protocol, subdomain, second-level domain, top-level domain and page path

## URL - Uniform Resource Locators

**A typical URL consists of the following parts**:

<https://www.example.com:443/path/to/page?query=123#fragment>

* Protocol (Scheme): http, https, ftp, etc.
  + Malicious URLs often use http instead of https to avoid encryption.
* Domain: www.example.com
  + Suspicious domains may have typos (e.g., g00gle.com).
* Port: 443 (Default for HTTPS)
  + Some attacks use non-standard ports.
* Path: /path/to/page
  + Look for encoded payloads or long, unusual paths.
* Query Parameters: ?query=123
  + Often used for SQL injection (id=1 OR 1=1).
* Fragment: #fragment
  + Not used for security checks, but phishing links may hide real destinations.

### **Features to Identify Malicious URLs**

To classify a URL as **malicious or benign**, consider extracting and analyzing the following features:

#### **1️⃣ Lexical Features (URL String-Based)**

* **Length of URL** – Longer URLs are more suspicious.
* **Number of special characters (@, -, \_, .)** – Phishing URLs often include multiple separators.
* **Number of digits in the URL** – Attackers use numeric domains (192.168.1.1).
* **Number of subdomains** – Example: login.facebook.secure.com (fake).
* **Presence of "@" in URL** – Redirects to another URL.
* **Presence of “-” in domain** – Example: secure-paypal.com (fake).
* **Presence of IP address instead of domain** – Example: http://192.168.1.1/login.

#### **2️⃣ Domain-Based Features**

* **Age of Domain** – New domains (< 6 months) are often malicious.
* **WHOIS Information** – Check if a domain is privately registered.
* **Reputation Score** – Check against databases like Google Safe Browsing.
* **Presence in Blacklists** – Malware domains are listed in threat intelligence sources.

#### **3️⃣ Content-Based Features**

* **Presence of Redirections** – Phishing sites often redirect multiple times.
* **Presence of Obfuscated JavaScript** – Look for encoded scripts (eval(base64\_decode(...))).
* **Number of iframes** – Hidden iframes are used for clickjacking.

#### **4️⃣ Behavioral Features**

* **Traffic Analytics** – Low or unusual traffic can indicate a fake site.
* **Abnormal Page Requests** – High request volume from different IPs.
* **Abnormal Referrer Information** – Sudden referrals from phishing emails.

#### **5️⃣ Machine Learning-Based Features**

You can train a model using extracted features to classify URLs as **malicious** or **benign**. Popular models include:

* Random Forest
* XGBoost
* Isolation Forest (for anomaly detection)
* Deep Learning (LSTMs for sequence-based classification)

E**xamples of malicious URLs**, categorized by attack type:

### **🔹 1. Phishing URLs (Fake Login Pages)**

Phishing sites trick users into entering credentials by imitating real websites.  
Example:  
🔴 **http://secure-paypal.com/login** (Fake PayPal login page)  
🔴 **http://facebook.verify-account.com/** (Fake Facebook verification page)  
🔴 **http://amazon-login.account-secure.com/** (Fake Amazon login)

📌 **Indicators of Phishing URLs:**  
✅ Looks similar to real domains but has **extra subdomains** (verify-account.com)  
✅ Uses **HTTP** instead of HTTPS  
✅ Contains **words like "secure", "verify", "login"** to gain trust

### **🔹 2. Malware-Hosting URLs (Drive-by Downloads)**

These URLs host malicious files that execute automatically when visited.  
Example:  
🔴 **http://freegames.download-now.com/setup.exe** (Trojan downloader)  
🔴 **http://example.com/get.php?file=virus.exe**  
🔴 **http://hacked-website.com/ads/script.js**

📌 **Indicators of Malware URLs:**  
✅ Ends with .exe, .bat, .zip, .js, .php (executables or scripts)  
✅ Hosted on unknown or hacked domains  
✅ Uses **"free", "crack", "download"** in the URL

### **🔹 3. Command and Control (C2) Server URLs**

Used by malware to communicate with a hacker’s server.  
Example:  
🔴 **http://123.45.67.89/connect.php?key=ABC123**  
🔴 **http://botnet-control.com/api/send\_command**  
🔴 **http://darkwebsite.com/payload?data=infected**

📌 **Indicators of C2 URLs:**  
✅ Often **use IP addresses** instead of domain names  
✅ Contain **strange query parameters** (?key=ABC123)  
✅ Hosted on **dark web or newly registered domains**

### **🔹 4. URL Shorteners & Redirections (Hidden Links)**

Attackers use URL shorteners to **hide** the real link.  
Example:  
🔴 **https://bit.ly/3Kfj8d9** → (Redirects to malicious site)  
🔴 **https://tinyurl.com/hackyourpc**  
🔴 **https://t.co/secure-login-paypal**

📌 **Indicators of Malicious Shortened URLs:**  
✅ Uses **URL shorteners** (bit.ly, tinyurl, goo.gl)  
✅ Hidden final destination  
✅ Often **spread via emails, SMS, or social media**

### **🔹 5. Exploit Kit URLs (Hacked Websites)**

Hackers inject malicious scripts into legitimate websites.  
Example:  
🔴 **http://hackedsite.com/wp-content/themes/exploit.js**  
🔴 **http://example.com/login.php?cmd=<script>evil\_code</script>**  
🔴 **http://legit-website.com/.well-known/favicon.ico** (Hiding malware in favicon files)

📌 **Indicators of Exploit URLs:**  
✅ Injects **JavaScript (.js)** or PHP backdoors  
✅ Uses **directory paths like /wp-content/** (WordPress attacks)  
✅ Often **uses iframe injections or obfuscated code**

### **🔹 6. Fake Tech Support URLs**

Pretends to be from Microsoft, Apple, or Google, showing fake virus warnings.  
Example:  
🔴 **http://microsoft-security-alert.com/your-pc-infected**  
🔴 **http://apple-support-warning.com/call-1800-123-4567**  
🔴 **http://fix-your-pc-now.com/error500**

📌 **Indicators of Fake Tech Support URLs:**  
✅ Uses **words like “support”, “alert”, “virus”**  
✅ Shows **popup warnings** when visited  
✅ Asks users to **call a fake support number**

### **🚀 How to Detect Malicious URLs?**

1️⃣ **Check Domain Age** – New domains (<6 months) are suspicious  
2️⃣ **Check Reputation** – Use Google Safe Browsing API  
3️⃣ **Look for Unusual Characters** – -, \_, @, %20 (URL encoding)  
4️⃣ **Verify HTTPS Usage** – Many phishing sites don’t use SSL  
5️⃣ **Use Machine Learning Models** – Train classifiers with URL features

Would you like a **Python script** to check if a URL is malicious? 🚀

# TLD -  top-level domain

Top-level domains come at the end of domain names. They are important for classifying domain names and essential for DNS lookups.

In the DNS hierarchy, a top-level domain (TLD) represents the first stop after the root zone.

Some other popular TLDs include ‘.org’, ‘.uk’, and ‘.edu’.

DNS resolvers start the search by communicating with the TLD server :

* **‘.com’** is intended for commercial businesses.
* **’.gov’** is for U.S. government entities.
* **’.uk’** is for domains from the United Kingdom.

### What are the different types of TLDs?

* **Generic TLDs:** Generic TLDs (gTLDs) encompass some of the more common domain names seen on the web, such as ‘.com’, ‘.net’, and ‘.org’.
* **Country-code TLDs:** Country-code TLDs (ccTLDs) are reserved for use by countries, sovereign states, and territories. Some examples are ‘.uk’, ‘.au’ (Australia), and ‘.jp’ (Japan
* **Sponsored TLDs:** These TLDs typically represent professional, ethnic, or geographical communities. Each sponsored TLD (sTLD) has a delegated sponsor that represents that community. For example, ‘.app’ is a TLD intended for the developer community. Similarly, ‘.gov’ is intended for use by the U.S. government.
* **Infrastructural TLDs:** This category only contains a single TLD: ‘.arpa’.
* **Reserved TLDs:** Some TLDs are on a reserved list, which means they are permanently unavailable for use. For example, ‘.localhost’ is reserved for local computer environments, and ‘.example’ is reserved for use in example demonstrations.