

COMP3310/6331 2020 – Tute/Lab #5

#	
1	<p>Exercises:</p> <p>With the release of assignment 2, you should (a) make sure you get your code working from the previous tute - check against the sample answer, and then (b) start to modify it as a http client. Use the tute time to get advice and/or review by the tutor and classmates. There are some 'telnet'-based exercises below, that may be helpful to debug your assignment code later on.</p> <p>You must also read (some of) RFC 1945 from the IETF. Focus on the protocol messages and errors mostly, and make sure you understand what you need to send, and what you may receive, in order to answer the assignment questions.</p>
2	<p>Review questions</p> <ol style="list-style-type: none"> What is the 'root' of the Domain Name System? i.e. what is at the very top? <ul style="list-style-type: none"> It's just "." - the dot, or fullstop. It's there to reference the root DNS servers (a-m.root-servers.org) Try to manually replicate a full iterative DNS lookup of www.canberratimes.com.au. using 'dig' and look for the NS records. You'll need to specify the nameserver as you go along using the '@' symbol. So pick one of the root servers, and then e.g. 'dig @f.root-servers.org ns au' which gives you all the .au nameservers. Note that if you do this repeatedly you'll get a different answer order. Then query the first of the .au nameservers for the com.au domain, and so on down the tree till you get to the www IP address. Where is the canberratimes hosted? What about their DNS entries? <ul style="list-style-type: none"> Looks like their DNS is hosted by Amazon (AWS-DNS) but their content is on EdgeKey, owned by Akamai. Let's look at some Fully Qualified Domain Names (FQDN) with different Top Level Domains (TLD). Consider: www.anu.edu.au. and www.gov. <ul style="list-style-type: none"> Which part is the 'generic TLD' (gTLD)? <ul style="list-style-type: none"> Edu and gov are the generic TLDs, though in AU we put the gTLD into the second-level Which part is the 'country code TLD' (ccTLD)? <ul style="list-style-type: none"> AU for the first, and none-specified for the second. Companies in the US don't use the .us ccTLD much at all, though many/most .com domains are not in the US. It's a very blurry use of the DNS, it applies to everybody except the US. Which part is the hostname? <ul style="list-style-type: none"> The 'www' in these cases. You can look up an A record for them in the DNS, they have a specific IP address (or set of addresses) so these are actual hosts you can connect to. Can you have both a 'gTLD' and a 'ccTLD'? <ul style="list-style-type: none"> No and yes. The gTLDs are meant to be the first (top) layer below '.' But many countries (outside of the US) have replicated them into their own 'top-level. It just happens to have their ccTLD after it. However,

this is not true of all countries. NZ, UK, JP use different 'generic' top-level (ok, second-level) domains to AU

4. In the case of cbs.anu.edu.au is the 'cbs' a hostname or a domain name, or both? How can you tell the difference? What's the benefit/downsides of this approach?
 - It can be a host AND a domain. There can be an A record (useful for its website) and it can have hostnames registered below it, e.g. mail.cbs.anu.edu.au, www.cbs.anu.edu.au and so on. It's convenient for keeping names shorter, especially for the website, and if the name is already quite long to remember. No real downsides, as long as you check before opening a connection.
5. In HTTP, how does the application handle the 'session' and 'presentation' layers?
 - Sessions are handled via Cookies (Set-Cookie: and Cookie: headers), or also via URL-referenced tokens - some 'query' field (after a '?') is passed to every page you go to, and it adds the same field to every URL in any page it sends back to you.

After here the answers depend on the website you choose. If you happen to pick a boring one, try a different site 😊

6. Use 'telnet' to replicate the steps from the lecture to connect to port 80:
 - telnet www.google.com 80 (or a different webserver of your choice)
 - Enter GET / HTTP/1.0 (and hit <return> twice)
 - Do you get redirected, or presented with a page of content?
 - If redirected to another host, use that server instead and start again
 - What do the headers tell you? Can you see Cache, Cookie, Accept type of headers? Do you see any size or timestamp headers? Keep in mind what happens with dynamically generated webpages.
 - Repeat with HEAD / HTTP/1.0 (and hit return twice)
 - Compare to what you got in the previous step - what's different?
 - Repeat and make a deliberate mistake in the command (e.g. HTTP/4.0, GAT instead of GET, or /comp33106331 instead of /). What kinds of errors do you get back? Can you get the server to return other errors?
7. Check your chosen webpage through webpagetest.org. How many different objects and object-types are involved to view just that page? How many sites? How quickly does it get everything? What's the biggest bottleneck?
8. Pick a hugely popular website in your part of the world, and run the above test. Can you tell if they are using a Content Distribution Network? If yes, which one? You can usually tell if a lot of their content comes from a different domain.