Paper selected: Great Firewall Paper

Answered by **Shyam sundar Ramamoorthy (shra3971)**

1. Why is the goal or motivation of this paper? In other words, why did the authors bother to write it and publish it?

   The goal is to find infrastructure of Great Firewall. Most of websites are blocked by GFW which is against the Freedom to Connect. The author provides results on circumvention protocols that gets blocked by use of active probing and finally describes how enhancement in circumvention protocol can eliminate active probing.

2. Does the paper state one or more hypothesis? What is the hypotheses of the paper? In what section(s) are they posed and in what sections are they answered?

   Hypothesis is "active probes originated mostly from different IP address in China which was triggered by real users (genuine ssh logins)". This is stated in Section2: Related work. Section4: Experiments provides favoring result towards this hypothesis. Section5.5 Fingerprinting Active probers shows a non-conclusive result that GFW conducts probes through a large distributed proxy network.

3. What kind of study is this (i.e. measurement, analysis, etc). It may be that the study involves more than one method. Provide the section numbers from the paper to justify your choice.

   Both measurement and analysis study in this paper.
   Measurement: Many experiments are conducted for identifying infrastructure of GFW. Section 4.1, 4.2 set measurement infrastructure.
   Analysis: Log analysis is done for finding probing period. Section 4.3

4. What are the parameters in the system? In other words, what are all the things that **could** be varied in this study? You obviously can't be exhaustive here, but be thoughtful and provide some examples.
   Again, for each parameter, provide the subsection number where it is first mentioned or discussed.

   IP address, Rate of probes, Location of probers are parameters. Geotagged location of probers, subnet of probers, at what time in a day more probes occur has to be determined. If more probes comes through a router/gateway, packets can be analyzed and dropped coming through that gateway.
   Section 5.2 provides info on time of probes,
   Section 5.3 for location & ip address of probers.

5. What are the factors in the study? This is what **was** varied in the study.
   Again, for each factor, provide the subsection number where it is first mentioned or discussed.

   Factors are types & payload of probes, topology of probers. Active probing being the major study, probes for finding different circumvention protocol is mentioned in Section 5.5.

6. What are the performance metrics used in the analysis?
   1. Again, for each performance metric, provide the subsection number where it is first mentioned or discussed.
   2. For each metric, state the type (rate/quantity/etc)
   3. For each metric, state if it is a LB/HB/NB metric
      Some metrics may not have a clear LB/HB metric; if so, indicated "unknown"0p9o]\

Metric	**Section specified in**	**Type of metric**	**LB/HB/NB**
Response time between connection and first probe	**Section 5.2**	**Rate**	**NB**
Number of probers	**Section 4.3**	**Quantity**	**LB**
TTL of syn packets (Bridges inbetween server & prober)	**Section 5.5**	**Distance**	**LB**
Counterprobes	**Section 4.4**	**Quantity**	**HB**

7. What is the goal or point of section 4.1 in the paper?

   - **This is an experiment to prevent bridges being discovered by probers.**
   - **To find if amazon EC2 servers are treated differently from other servers.**
   - **To determine how probing differs for Tor, obfs2, obfs3**
   - **To find prober's ISP network.**

8. Figure 4 and Figure 8 both show time-varying activity and both have "gaps" between periods of activity. Are the gaps related to one another? Justify your answer using information from the paper.

   Fig 4 shows successful connection attempts of tor clients to tor bridges without getting blocked and Fig8 show probe interval for various protocol.
   Both gaps are related to one another because probes block connection of client & bridge. Clients from Unicom server is less successful than CERNET clients because Unicom operate on same link as GFW and CERNET is one hop closer to circumvent servers.

9. Figure 8 shows a time plot of probe types *vs.* volume for different types. Explain at least two examples of information the authors hoped to convey by showing this as a time series plot rather than a simple summary (*e.g.* boxplot or average).

   Author compares identification of circumvention protocols by active probes.eg: Obfs2 is easily detected as first few bytes of TCP connection can be used to decrypt data. Author also compares probe data, data used for circumvention and ordinary data requests.
   Also compares probes rate for different ports of server.

10. Assume you were the person reviewing the paper. What changes would you want to be made in the data analysis and comparisons?

    More servers like EC2 has to be analyzed. To see if most of the probers are from same virtual network. Private IP can be analyzed which connects through distributed proxies.

11. The paper has several conclusions or lessons. Which lesson do you think is "most overblown" or least justified by the data presented in the paper. This may turn on the phrasing of the lesson or the justification presented.

    Paper states its highly unlikely that system used packet injection for probing. There can be NAT boxes may hijack public ip address and assigned it to probers in private network and there is a chance that NAT boxes delay the ACK packets to be sent by processing SYN_ACK from circumvention servers.