Developing techniques to identify attack traffic in Tor

Rachee Singh and Rishab Nithyanand

April 26, 2019

We are seeking feedback on our proposal for identifying attack traffic within the Tor network.

1. What are you trying to learn, and why is that useful for the world? That is, what are the hoped-for benefits of your experiment?

Background. Recent studies have shown that malicious actors leverage the Tor network to launch a wide range of attacks including brute-force logins, port scanning, *etc.* Evidence from prior research has shown that, as a consequence of these attacks, several IP blacklists currently discriminate against Tor exit relays (4). In general, these attacks have led to the differential treatment of Tor exits by Web servers (2).

Goal. Our goal is to develop techniques to identify when the Tor network is being used for propagation of such attacks. As a first step, we will conduct a measurement study on Tor exit relays that we own to identify when they are being used to launch attacks.

Why is this important. Equipped with these measurements, we plan to investigate techniques to identify and differentiate (by blocking or rate-limiting) attack traffic within the Tor network. Our hope is that the conducted research will result in solutions that reduce the incidence of abuse and attacks via the Tor network and consequently motivate content providers to not discriminate against Tor exits, thereby increasing the utility of the Tor network.

2. What exactly is your plan? That is, what are the steps of your experiment, what will you collect, how will you keep it safe, and so on.

Our current plan is to leverage existing techniques (3) for identifying common types of attack traffic (*e.g.,* SSH brute force logins, *etc*). To apply these techniques, we plan to measure traffic characteristics in the following way:

1. **Host honeypots.** We plan to host a number of honeypot servers which are reachable over commonly accessed ports (Port 22 for SSH, SMTP port, *etc*). We will log all traffic received on this server. With the data gathered by the honeypots, we can identify the incidence rates of different types of attacks from Tor and non-Tor users.

2. ~~**Host exit relays in the Tor network**~~**Pilot study.** We ~~plan to set up several exit relays with different configurations of bandwidth and exit policy for the duration of this study (3-6 months).~~

3. ~~**Attempt to attack our own server.** In addition to leaving our servers vulnerable to attacks by others, we will also launch our own attacks through the Tor network. To do this , we will use a Tor client to establish circuits using our exit relays. Using these connections, we will launch the most common implementations of well-known attacks (for example, brute-force login attempts) on our own honeypot servers.~~

4. ~~**Collect information on our exits.** On each of~~ will deploy an exit relay and acquire honeypot servers. We will restrict the exit policy of our relay to our ~~exit relays~~honeypot servers' IP addresses and specific port numbers to ensure other Tor users do not use our exit. We will launch attacks on our honeypots, while routing traffic through this exit relay. This attack traffic will be blended with traffic from a simulated user model (1). For different sampling granularities (all packets, one-in-k packets, one-in-k packets per circuit) we will evaluate the accuracy of our techniques in detecting attack traffic. The coarsest sampling granularity which offers the highest accuracy of attack detection will then be deployed in the wild.

   We will collect traffic on the egress of our exit.

5. **Real world deployment.** At the appropriate sampling granularity derived from our pilot study, we will ~~collect the following information from packets corresponding to each Tor circuit~~log the following data on exit relays open to all Tor users:

- *MAC of the destination IP address.* ~~Note that we do not plan to store IP addresses of destinations~~Instead of the destination IP address, we will log only the MAC of the destination IP. We will ~~use a MAC whose key will never be written to disk~~not store the key of the MAC. Our techniques are destination-agnostic and only require information about the number of simultaneous connections to each destination IP rather than knowledge of the actual destination IP. Therefore, MAC'd destination IPs are sufficient for our study.
- *Destination port.* We plan to log the destination port of the packets to differentiate between attack types.
- *Packet size.* We plan to log the size of ~~each packet~~ sampled packets (number of bytes). This will help us identify anomalous increase in traffic volume towards particular destinations.
- *Middle-relay IP address and port.* We will collect IP addresses and port numbers of the middle-relays propagating the traffic ~~seen at~~ to our exits. Since we will not ~~be running~~ run any of our relays as guards, we ~~will not be able to~~ cannot perform correlations to de-anonymize users. ~~Also~~Additionally, since we will not ~~be logging~~ log actual destination IP addresses, even if third-parties running guard relays ~~who are able to gain (unauthorized)~~ were to gain unauthorized access to our logs~~will not be able~~ , they will be unable to de-anonymize ~~the Tor usersusing our exits~~Tor users.
- *Tor circuit ID.* We plan to log the circuit ID of ~~each packet~~ sampled packets exiting our relay. This will help us identify which packets belong to the same stream (and consequently belong to the same attack stream).
- *Sets of co-occurring circuits.* This will help us identify when the Tor network is potentially being used for distributed attacks. This also allows us to perform our analysis without logging timestamps for each packet.

While the relays are open, we will continue to launch our own attacks through these exit relays and to our own honeypots. Therefore the gathered data will contain sampled traces of real Tor traffic and our own attack traffic.

6. **Transfer logs to analysis server.** We plan to transfer the collected logs to an external server for analysis at the end of each hour. Using these logs, we will apply the techniques outlined in prior work (3) to flag traffic with anomalous characteristics that are indicative of attacks.

3. What attacks or risks might be introduced or assisted because of your actions or your data sets, and how well do you resolve each of them?

~~Our~~ Collected datasets will consist of one ¡$ID_{circuit}$, list of co-occurring circuits: $[ID_{circuit_1}, \ldots, ID_{circuit_n}]$, $IP_{middlerelay}$, $port_{middlerelay}$, $MAC_k(IP_{destination})$, $port_{destination}$¿ tuple per ~~packet observed by~~ sampled packet on our exit relays. We take several actions to prevent the user de-anonymization ~~,~~ due to the above data logging~~, of users of our exit relays.~~ :

Preventing end-to-end traffic correlation de-anonymization. We only store ~~cryptographic hashes~~ the MAC of destination IP addresses. This prevents the direct de-anonymization of users by entities able to observe our logs and user traffic entering the Tor network (*e.g.,* guard relays and user ISPs).

Preventing timing attacks. We do not log timestamps of packets. This will prevent user de-anonymization due to timing correlation attacks by entities able to observe our logs. A new risk might involve an adversary forcing creation of new circuits that can be observed to co-occur with others in our logs. Such an adversary might still be able to perform timing correlations if they are able to observe our logged data. To address this issue we reduce the attack window through our third attack mitigation approach: at the end of each hour, we will transfer our logs onto well protected servers at the University of Iowa where they will be encrypted and password protected.

4. Walk us through why the benefits from item 1 outweigh the remaining risks from item 3: why is this plan worthwhile despite the remaining risks?

Comparable attack surface with current Tor. To our knowledge, the information we propose to collect on exit relays does not expose Tor users to any currently known attacks after our mitigations are applied. We believe that the attack surface made available by our logging and mitigations: (1) generate circuits going through our exit relays so that they co-occur with the victims circuits going through our relays and (2) compromise our exit relays to obtain these logs within the hour of the attack (or) compromise our password-protected encrypted database at UIowa, is comparable to Tor's existing attack surface where an adversary might compromise an exit to observe and correlate Tor user traffic.

High potential payoffs. At a high-level we expect that our research will result in the development of methods to improve the utility of the Tor network. This will be possible due to the potential reduction in the incidence of attacks performed via the Tor network. Such a reduction in attack traffic

from the Tor network can lead to an improvement in the reputation of exit relays on IP blacklists. This in turn might result in network operators and content providers unblocking Tor exit relays.

References

[1] Aaron Johnson, Chris Wacek, Rob Jansen, Micah Sherr, and Paul Syverson. Users get routed: Traffic correlation on tor by realistic adversaries. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, CCS '13, pages 337–348, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2477-9. doi: 10.1145/2508859.2516651. URL http://doi.acm.org/10.1145/2508859.2516651.

[2] Sheharbano Khattak, David Fifield, Sadia Afroz, Mobin Javed, Srikanth Sundaresan, Vern Paxson, Steven J Murdoch, and Damon McCoy. Do you see what I see? differential treatment of anonymous users. Internet Society, 2016.

[3] Rui Miao, Rahul Potharaju, Minlan Yu, and Navendu Jain. The dark menace: Characterizing network-based attacks in the cloud. In *Proceedings of the 2015 Internet Measurement Conference*, pages 169–182. ACM, 2015.

[4] Rachee Singh, Rishab Nithyanand, Sadia Afroz, Paul Pearce, Michael Carl Tschantz, Phillipa Gill, and Vern Paxson. Characterizing the Nature and Dynamics of Tor Exit Blocking. In *26th USENIX Security Symposium (USENIX Security 17)*, pages 325–341, Vancouver, BC, 2017. USENIX Association. ISBN 978-1-931971-40-9. URL https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/si