Dear Dr. Rodriguez,

We thank you and the referees for reviewing our paper. Below we provide detailed answers to the reviewers’ comments. Our responses are marked by (*). We highlight the changes that were made in the previous version with blue and changes made in this round with red.

Best regards,

Ionut Trestian
Supranamaya Ranjan
Aleksandar Kuzmanovic
Antonio Nucci

Reviewer(s)’ Comments to Author:

Reviewer: 1
Comments to the Author

The paper benefits greatly with the revisions with regard to
a) toning down the claims about UEP’s utility in isolation and a better explanation of how UEP complements existing traffic classification approaches
b) better evaluation of UEP’s accuracy by comparing with a payload signature based tool

Below are a list of things that need fixing:

-- The captions of almost all figures need more detail. You can’t have “traffic destinations” (Fig. 3) and “IP addresses” (Fig. 4) as figure captions! The caption of a graph needs to outline the experiment that produced the graph so that the reader can interpret the graph without having to go through all the text. The following URL has a few guidelines that are worth following when captioning figures: [http://www.physics.ohio-state.edu/~wilkins/writing/Handouts/fig-captions.html](http://www.physics.ohio-state.edu/~wilkins/writing/Handouts/fig-captions.html)

(*) We agree with the reviewer that it would be useful for the captions to contain more information and we have modified them accordingly

---

-- For the results in Table V and VI, it would be valuable to have the absolute number of flows in addition to the percentages cited. For example, for streaming, mail, and FTP traffic, though the absolute difference in percentages between UEP and signature-based classification is small, there is a significant relative difference. It’s hard to say how significant the relative difference is without access to the absolute number of flows.

---
The authors agree with the suggestion of the reviewer. We have added the absolute number of flows in Table 5. Table 6 already contains the absolute number of endpoints.

---

The study on staleness of information discovered by UEP (second column of page 13) is plain incorrect. What you study is the staleness of the webpages from which you extract information about IP addresses. This does not tell you anything about whether the information you extract from that page is still valid. For example, on an online forum, the page corresponding to a particular thread may be updated every day as new users post to that thread. However, information you initially extracted from this page, e.g., that the IP address of the first poster on this thread is associated with say gaming, may no longer be valid.

To perform a valid study of staleness, you need ground truth, for which you need a trace spanning a long interval. If you are unable to gather such a trace, please tone down your claims about the staleness of data yielded by UEP.

---

We have toned down the claims about the staleness of the information present on the Web and we have added the following line: “Although we cannot validate that the information about a particular IP address is accurate this measure still shows a fairly high freshness of the websites that display IP address information.”

---

In the second column on page 8, it should be "only by the signature based traffic classification (S-U) and only by our approach (U-S)"

---

We have addressed this.

---

Reviewer: 2

Comments to the Author

I am reviewer 3 from the previous set of reviewers. I like to state that all my major concerns have been addressed, though a few concerns (mostly editorial) remain. First let me mention a few high-level concerns.

-I believe the confusion regarding the word "unconstrained" is due to people confusing "different" with "better". BLINC etc. used network flow data by design. Your solution uses
information that can be searched via the WWW to achieve the same task, by design. Your solution is different, but that doesn’t automatically make it better. Just because your solution is “unconstrained” doesn’t make your solution better. Although I do realize you never adopted such a stance, but unfortunately it seems to come out that way.

(*) The authors understand the confusion given by the use of the word unconstrained. We never intended to suggest that “unconstrained” might mean “better”. In order to address this we have added the following sentences in the introduction: “Hence our approach is different by design (not necessarily better) from other traffic classification approaches (e.g. BLINC). We compare these approaches for given networks later in the paper.”

---

Regarding traffic locality, if you have the information, it might be nice to show what percentage of the total traffic in areas with high locality is P2P. For instance in China, where you show high locality, is most of the traffic P2P or other apps? This would be very interesting for multiple reasons. If a high percentage of total traffic is local, and if a high percentage of the traffic that is local is not P2P, this can have fundamental consequences in how traffic patterns in the Internet are modeled now. For instance the gravity model to model traffic matrices will not hold under locality.

(*) The authors would like to perform such a study, however given the fact that all used approaches didn’t manage to classify a given percent of the traffic, making absolute claims about this would be hazardous.

---

Some minor comments:

- Pg 3, Col 2 Line 9: Sentence incomplete!

(*) We have addressed this.

---

- Pg 3, Col 2 Line 17: “.. extract the ones which..” -> ‘which’ to ‘that’. There are other places also where which should be changed to that – please go thru the text carefully.

(*) We have addressed this.
---

Pg 8, Col1, Line 39: "The first reason.." -> Rephrase, very awkward sentence construction.

---

(*) We have addressed this. The phrase now reads: “The UEP approach is online capable because of its ability to classify traffic based on a single observed packet for which one of the endpoints is revealed (e.g., a web server). Furthermore, there is a huge bias of traffic destinations ...”

---

Pg 8 Col1, Line 54: "most popular 5%.." -> popular where? 5% of the most popular IP addresses in the whole known universe? Please be specific.

---

(*) We have addressed this. “most popular 5% from the traces”

---

Reviewer: 3

Comments to the Author

The authors seem to have addressed some of the issues raised by the reviewers.

However a number of issues still remain:

The limitations of the UEP methodology are never clearly identified and discussed properly. For example, the authors discuss the need/absence of packet traces in two paragraphs in the introduction as a limitation for previous work. But to provide sufficient classification for a number of applications, UEP also needs a lot of “side” information such as crawling p2p networks. This is never discussed in the introduction.

Thus, the authors should clearly state in the introduction (since most of the focus is towards application classification), that googling the Internet alone can only provide limited information about certain applications such as p2p networks, and significant effort might be needed to overcome such limitations (e.g., crawling a p2p network). The authors only vaguely allude to this fact and this can lead to misconceptions when reading the paper.

---

(*) We agree with the reviewer and we have incorporated the following paragraph in the introduction: “Still not all information is available on the Web, hence results may be improved by using additional sources of information, some of which come at a high cost (e.g. joining and crawling a p2p network).”
Also, when discussing comparisons with other classification methodologies, they should make clear that UEP outperforms other methodologies in the particular cases and networks examined (which is again only alluded in later sections).

The authors should devote a couple of paragraphs clearly discussing several choices made throughout the paper and the limitations that these might entail (This might be towards the end of the paper in the discussion section or a separate limitations section). These include the assumption of neighboring IPs, the timescale issue which is only touched by the reviewers, the fact that only particular networks are examined, browsing through anonymization proxies or networks etc. This is needed so readers can clearly identify the strengths and weaknesses of UEP.

(*)Regarding the assumption of neighboring IP addresses this is based on the fact that IP address assignment on the Internet at different points in the assignment hierarchy is done on a block basis. Indeed, assignment agencies assign specific contiguous blocks to ISPs and also higher tier ISPs assign contiguous blocks to lower tier ISPs. It is also easier to specify blocks in access control lists, filters than to specify individual hosts. We have added the following paragraphs in the paper to account for this: “Indeed, IP address assignment on the Internet at different points in the assignment hierarchy except the end user is done on a block basis. Also, to ease network management, network administrators typically assign contiguous IP addresses to hosts in the same network”

We have added paragraphs in the introduction referring that our comparison holds for the particular networks examined:

“Hence our approach is different by design (not necessarily better) from other traffic classification approaches (e.g.BLINC). We compare these approaches for given networks later in the paper”

“It should be noted that the examined networks belong to tier 1 ISPs which is an unfriendly environment for one of the compared approaches [25].”

We have toned down the claims about the timescales and staleness of the information present on the Web and we have added the following line: “Although we cannot validate that the information about a particular IP address is accurate this measure still shows a fairly high freshness of the websites that display IP address information.”

Regarding browsing through anonymization proxies our paper didn’t draw any conclusions regarding the particular websites that people access
through these proxies but only with regard to the browsing behavior which is still captured.

---

Especially for the timescale issue, the data provided are not convincing. The authors show that 30%–40% (depending on how one interprets the 88% and 71% numbers, 88%*71%=62%) of the top 300 websites do contain stale information, which is a significant fraction and might lead to misclassifications or false findings. The authors should make clear that this might be a limitation (it would be interesting to examine this in future work).

---

(*)We have toned down the claims (see comment from previous reviewer) about the staleness of the information present on the Web and we have added the following line: “Although we cannot validate that the information about a particular IP address is accurate this measure still shows a fairly high freshness of the websites that display IP address information.”

---

Similarly for p2p networks. The authors added text mentioning that the entry point of p2p will be known publicly. They also need to make clear that these entry points are however only a small fraction of the whole p2p network (or provide evidence against this).

---

(*)We agree with the reviewer and have added the following phrase to acknowledge this: “The number of such endpoints in a p2p network is relatively small however whenever a client wants to retrieve a file he typically goes through such an access point.”

---

The authors added ground truth analysis comparing with a signature based classifier. This is nice but on the other hand, it clearly highlights the points above, namely that for applications besides web where information is public, UEP fails with no other “side” information. The authors should clearly identify this, namely that false negatives compared to signature based classification are roughly 50% for p2p and streaming applications, 23% for email and 36% for ftp. Since results are presented in flows, one would expect the overall classification to be significantly low in terms of bytes (taking into account that p2p and streaming would contribute most in terms of bytes). This is only vaguely discussed by the authors.

---

(*)We agree with the reviewer and we have added this in the text. “(e.g. for mail 22,365 flows for UEP compared to 30,426 flows for signature based and for FTP 1,338 flows for UEP compared to 2,278
flows for signature based).” We have also added the number of flows in Table 5 to illustrate this.

Our paper was never targeted at identifying traffic volumes and consequently we didn’t make claims regarding this.

---