User survey regarding the needs of network researchers in trace-anonymization tools

Jihwang Yeo, Keren Tan, David Kotz
Dartmouth Computer Science Technical Report TR2009-658

November 23, 2009

Abstract

To understand the needs of network researchers in an anonymization tool, we conducted a survey on the network researchers. We invited network researchers world-wide to the survey by sending invitation emails to well-known mailing lists whose subscribers may be interested in network research with collecting, sharing and sanitizing network traces.

1 Survey set-up

Network traces, which record the activity of network users, are an important tool in computer networking research. It remains a difficult challenge to capture such traces, primarily because it is difficult to obtain permission from the network operator. It is even more difficult to share network traces with others, because of privacy concerns.

We hope increase network-trace sharing by making it safer and easier to "sanitize" network traces, that is, to remove sensitive identifiable information. Sanitization always involves a challenging trade-off between sanitization effectiveness (providing anonymity for network users and secrecy for network operational information) and research usefulness (since only the information retained can be used by the researcher).

We set out to survey network researchers to determine what experience they had with collecting, sanitizing, or using network traces. We asked about 30 questions regarding collecting, sharing and sanitizing network traces, using sanitization tools, and evaluating sanitization results. The survey takes 5-15 minutes but the survey participant can stop earlier if he or she wishes. In the end, 108 people participated in the survey. We conducted our survey during the period from May 21, 2009 to October 27, 2009.

2 Survey results

According to their responses, 84% of the participants have collected network traces and 88% of the collected traces contain some sensitive information. More than 83% of the participants agree that protecting personal information is a major concern when sharing the traces with others and this issue is considered to be an obstacle. About 55% of the participants who have collected their own traces, have experience of sanitizing the traces. Among them, only 39% used third-party tools while about 71% of people used their own software. Regarding the verification of sanitization result, 84% of 38 responders answered that they did not use any quantitative metrics to measure the strength of the sanitization. Also, 76% of the responders thought that the sanitization they applied will limit the usefulness of network traces. About half of 86 participants have never heard or used any sanitization tools. The tools that participants reported to have heard or used for sanitization, include TCPdpriv, TCPurify, Crypto-PAn, AnonTool, tcpmkpub, FLAIM, CANINE, CoralReef, PktAnon, Tcpdump Anonymizer, traceanon, WDCap, in the order of most uses.

Figures 1–20 show the details of the survey results, including the actual questions, response statistics and summary of the responses.

3 Summary and conclusion

This information from the survey feedback will help us to refine our ideas about expressing sanitization goals and research usefulness goals, about new anonymization methods, and about new metrics. The information will also help us to develop the NetSANI framework.[1]

Acknowledgements

This survey results from a research program in the Institute for Security, Technology, and Society (ISTS), supported by the U.S. Department of Homeland Security under Grant Award Number 2006-CS-001-000001, and by the NetSANI project at Dartmouth College (funded by Award CNS-0831409 from the National Science Foundation).

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security or the National Science Foundation.

[1] http://www.cs.dartmouth.edu/~dfk/papers/index.html#netsani

Have you ever collected network traces for your research?

Answer Options	Response Percent	Response Count
Yes	83.5%	96
No	16.5%	19
answered question		**115**
skipped question		**0**

Figure 1: (Question 1) 84% of the participants have collected network traces

What kinds of network traces have you collected? (check all that apply)

Answer Options	Response Percent	Response Count
NetFlow log	23.1%	21
Pcap trace	74.7%	68
NCSA Unified Format	0.0%	0
IEEE 802.11 (such as Prism or Radiotap header, or raw	39.6%	36
Bluetooth	7.7%	7
Authentication log	8.8%	8
Syslog	24.2%	22
NONE	2.2%	2
Other (please specify)	22.0%	20
answered question		**91**
skipped question		**24**

Figure 2: (Question 2) 75% of the participants who have collected network traces, collected pcap traces.

Sometimes network traces contain sensitive information - identifiable information linked to a network user (such as IP address and MAC address) or information of proprietary or operational significance (such as the location or structure of the network, or identity of important servers). Did the traces you collected contain any sensitive information?		
Answer Options	**Response Percent**	**Response Count**
Yes	87.8%	79
No	12.2%	11
	answered question	**90**
	skipped question	**25**

Figure 3: (Question 3) 88% of the participants who have collected network traces, responded that the collected traces contain some sensitive information.

Have you shared your traces with other researchers outside your organization?		
Answer Options	**Response Percent**	**Response Count**
Yes	43.2%	35
No	56.8%	46
	answered question	**81**
	skipped question	**34**

Figure 4: (Question 5) 43% of the participants who have collected network traces, shared their traces with other researchers.

Do you intend to share the traces with other researchers outside your organization?		
Answer Options	**Response Percent**	**Response Count**
Yes	65.2%	43
No	34.8%	23
	answered question	**66**
	skipped question	**49**

Figure 5: (Question 7) 65% of the participants who have collected network traces, intend to share the traces with other researchers outside their organizations.

Why you do not want to share the collected traces with other researchers outside your organizations? (check all that apply)		
Answer Options	**Response Percent**	**Response Count**
Protecting personally identifiable information	76.2%	16
Protecting proprietary or operational information	28.6%	6
Competitive issues	19.0%	4
Legal issues	61.9%	13
Other (please specify)	28.6%	6
	answered question	**21**
	skipped question	**94**

Figure 6: (Question 8) 76% of the participants who have collected network traces responded that they do not want to share the traces with other researchers outside their organization, because they need to protect personally identifiable information.

When you plan to share the traces, what will be your major concern(s)? (check all that apply)		
Answer Options	**Response Percent**	**Response Count**
Protecting personally identifiable information	83.7%	36
Protecting proprietary or operational information	62.8%	27
Competitive issue	25.6%	11
Legal issues	60.5%	26
Other (please specify)	11.6%	5
	answered question	**43**
	skipped question	**72**

Figure 7: (Question 9) 84% of the participants who have collected network traces responded that their major concern in sharing the traces with other researchers outside their organization is protecting personally identifiable information.

Generally, "sanitizing" network traces means processing the traces to remove the sensitive information or reduce the degree of its sensitivity. Did you sanitize the network trace?		
Answer Options	**Response Percent**	**Response Count**
Yes	54.5%	42
No	45.5%	35
	answered question	**77**
	skipped question	**38**

Figure 8: (Question 10) Only 55% of the participants who have collected network traces sanitized the traces.

How do you sanitize network traces? (check all that apply)		
Answer Options	**Response Percent**	**Response Count**
Manually editing the traces	12.2%	5
Using your own software	70.7%	29
Using third-party tools	34.1%	14
Other (please specify)	7.3%	3
	answered question	**41**
	skipped question	**74**

Figure 9: (Question 11) 71% of the participants who have sanitized their traces used their own software. 34% of the participants who have sanitized their traces used third-party tools.

How satisfied are you with the sanitization tool(s) you used							
Answer Options	**(1) Least satisfied**	**(2)**	**(3)**	**(4)**	**(5) Most satisfied**	**Rating Average**	**Response Count**
Satisfaction degree	5	9	8	9	9	3.20	40
					answered question		**40**
					skipped question		**75**

Figure 10: (Question 13) On average, the participants who have sanitized are satisfied with their sanitization tools; score 3.2 out of 5.

After sanitizing network traces, have you verified that your sanitization result is correct?		
Answer Options	**Response Percent**	**Response Count**
Yes	76.3%	29
No	23.7%	9
	answered question	**38**
	skipped question	**77**

Figure 11: (Question 15) 76% of the participants who have sanitized network traces have verified that their sanitization result is correct.

How difficult was it to verify the sanitization result?							
Answer Options	**(1) Very easy**	**(2)**	**(3)**	**(4)**	**(5) Very difficult**	**Rating Average**	**Response Count**
Difficulty degree	5	6	9	5	4	2.90	29
						answered question	**29**
						skipped question	**86**

Figure 12: (Question 16) On average, the participants who verified their sanitization result, found it difficult to do so; score of 2.9 out of 5.0.

Do you have any quantitative metric to measure the strength of the sanitization (such a metric indicates how well the sanitization removes or reduces sensitive information)?		
Answer Options	**Response Percent**	**Response Count**
Yes	15.8%	6
No	84.2%	32
	answered question	**38**
	skipped question	**77**

Figure 13: (Question 17) 84% of the participants who verified their sanitization result did not use any quantitative metric to measure the strength of the sanitization.

Do you think the sanitization you applied will limit the usefulness of network traces?		
Answer Options	**Response Percent**	**Response Count**
Yes	76.3%	29
No	23.7%	9
	answered question	**38**
	skipped question	**77**

Figure 14: (Question 19) 76% of the participants who verified their sanitization result agree that the sanitization they applied limits the usefulness of the network traces.

Do you have any quantitative metric to measure how much the sanitization actually reduced the usefulness of your network trace?		
Answer Options	**Response Percent**	**Response Count**
Yes	7.9%	3
No	92.1%	35
	answered question	**38**
	skipped question	**77**

Figure 15: (Question 20) 92% of the participants who have verified their sanitization result did not use any quantitative metric to measure how much the sanitization actually reduced the usefulness of their network trace.

Have you ever used network traces for your research?		
Answer Options	**Response Percent**	**Response Count**
Yes	61.1%	11
No	38.9%	7
	answered question	**18**
	skipped question	**97**

Figure 16: (Question 22) 61% of 18 respondents have used network traces for their research.

How did you obtain the network traces? (check all that apply)		
Answer Options	**Response Percent**	**Response Count**
Direct request to data authors	54.5%	6
Downloading from online network data archives (like	100.0%	11
Collecting your own traces	0.0%	0
	answered question	**11**
	skipped question	**104**

Figure 17: (Question 23) All of 11 respondents have obtained the network traces by downloading from on-line network data archives.

What kinds of network traces have you used? (check all that apply)		
Answer Options	**Response Percent**	**Response Count**
NetFlow log	18.2%	2
Pcap trace	27.3%	3
NCSA Unified Format	0.0%	0
IEEE 802.11 (such as Prism or Radiotap header, or raw	72.7%	8
Bluetooth	27.3%	3
Authentication log	9.1%	1
Syslog	27.3%	3
Other (please specify)	18.2%	2
answered question		**11**
skipped question		**104**

Figure 18: (Question 24) 72% of 11 respondents have used IEEE 802.11 network traces. 27% of the respondents have used pcap traces or syslog.

Sometimes network traces contain sensitive information - identifiable information linked to a network user (such as IP address and MAC address) or		
Answer Options	**Response Percent**	**Response Count**
Yes	45.5%	5
No	54.5%	6
answered question		**11**
skipped question		**104**

Figure 19: (Question 25) 55% of 11 respondents noticed some sensitive information in the traces they collected.

Have you heard or used the following sanitization tools? (check all that apply)		
Answer Options	**Response Percent**	**Response Count**
Crypto-PAn	22.1%	19
TCPurify	26.7%	23
TCPdpriv	27.9%	24
CANINE	9.3%	8
AnonTool	22.1%	19
tcpmkpub	17.4%	15
FLAIM	15.1%	13
NONE	48.8%	42
Other (please specify)	10.5%	9
answered question		**86**
skipped question		**29**

Figure 20: (Question 27) 49% of 86 respondents have never heard of, nor used, any of these sanitization tools