The Tracking of Users’ Unintentionally Shared Information by Social Network Sites

Rath Kanha Sar
Master of Information Technology
(Charles Sturt University, Australia)
Bachelor of Computer Science and Engineering
Bachelor of Education
(Royal University of Phnom Penh, Cambodia)

A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy
Submitted March 2014

School of Computing and Mathematics
Faculty of Business
Charles Sturt University
Australia
Abstract

Online tracking has been around ever since HTTP cookies were invented in the 1990s. The tracking technologies are becoming more sophisticated, and large amounts of online users’ information can be collected rapidly by numerous first and third party sites. This research aims to explore the privacy issues associated with the tracking of users’ information by major social network sites (SNSs) such as Facebook, Twitter and Google Plus, from the perspective of online users and from the moral perspective of a privacy framework.

The first stage of this research, which involves an exploratory study and experiments with 20 research participants from Cambodia, investigates the collection of online users’ information in HTTP headers. Among other outcomes, the experiment results indicated that online users’ information is being tracked by numerous third party sites including advertisers, data aggregators and major SNSs like Facebook, Twitter, and Google Plus. The rest of this research focuses on tracking by SNSs owing to the fact that more than one billion users are actively using SNSs. In addition, SNSs are not only acquiring large quantities of personal information voluntarily supplied by users (intentionally shared information), but also details about their browsing activities around other sites (unintentionally shared information). The combination of these types of information reveals much about a person’s life.

The second stage of this research gauges research participants’ awareness about online privacy and tracking by SNSs based on their browsing activities in the first stage of the research. It is influenced by ethnography and employs open-ended structured interviews in order to elicit the views of research participants. The outcomes of the interviews suggested that all participants valued privacy, either online or offline. It also revealed that 19/20 participants from Cambodia were not aware of online tracking by third party sites at all, nor did they read privacy policies. They perceived online tracking to be tracking by other Internet users who glean information through SNSs like Facebook. They expressed anger and concern about privacy loss when they learned that details about their browsing activities were being tracked by SNSs. They also stated that SNSs should have sought their permission before tracking their activities.

Last but not least, the third stage of this research analysed the practice of tracking users’ un-
intentionally shared information by SNSs from a moral perspective. It employed Contextual Integrity (CI) as a framework by applying its prescribed norms (norms of appropriateness and norms of distribution) as well as its decision heuristic. Analysis conducted by applying the CI framework suggested that SNSs violated online users’ right to privacy because they failed to obtain online users’ informed consent or permission before moving their information across different contexts (e.g., collecting users’ information from sites outside SNS boundaries). The analysis through the lens of the CI framework appears to be consistent with the findings from the perspectives of research participants in Cambodia: users’ informed consent or permission is needed before SNSs can collect their information across different sites.

This research did not only allow research participants to observe the sharing of their information in the HTTP headers, but also gauged their perceptions and awareness of the practice of online tracking. It is also the first ICT work that involved research participants from Cambodia. Unlike the existing literature that tested a large number of cases, information sharing examined in this research is based on smaller number of cases, but includes real life online users’ browsing activities. In addition, this is also the first research that involved analysing the impacts of tracking by SNSs from not only the perspective of online users, but also from a moral perspective. The findings from both perspectives are consistent. Hence, tracking of users’ unintentionally shared information by SNSs appears to be wrong. The researcher hopes that future work will investigate this practice from the perspective of other stakeholders such as SNSs, advertisers and data aggregators. For example, future research could focus on what happens to users’ unintentionally shared information once collected. This will enable decisions to be made about whether or not this practice is harmful to users.
List of publications


Acknowledgements

With the completion of my Doctor of Philosophy (PhD), the real value of this work has been not so much in the content, as in the many people who have become part of my life and PhD experience.

I would like to thank my principal supervisor, Associate Professor Yeslam Al-Saggaf, for his continuous support of my PhD, his frequent reviews of various drafts of my thesis, his advice, and his extra patience, through rain, hail or shine. I thank Professor Eddie Oczkowski, for his help, support and understanding with the change in my supervision team and research plan halfway through my journey. I have also appreciated the feedback and comments on the drafts of my thesis, provided by my associate supervisor, Dr Md Zahidul Islam.

My thanks also go to National ICT Australia (NICTA) and the University of Sydney for providing me with a scholarship and the opportunity to undertake a summer research project prior to my PhD journey. I would also like to thank Dr Sebastien Ardon and Dr Anirban Mahanti for their guidance, knowledge, motivation and time in supervising the project.

To Mike and Rozzie Dunn for all their help and encouragement during my stay in Wagga Wagga. I would also like to thank Maureen Engler for her helps and motivation in my applications both for the summer research scholarship at NICTA and PhD scholarship at Charles Sturt University (CSU). I thank my relatives, the Seng family for making me feel so welcome during my stay with them in Sydney during my summer research project.

Also, I would like to express my gratitude to the participants in Cambodia who were so generous with their time and trust. They shared their personal thoughts and demonstrated great interest in the topic of online privacy. I thank my dearest sisters, Rath Amara Sar and Malis Pich Sar, who helped to organise the workshop, book the venue and print the research materials for the workshop and interviews.

To my friends, Sophal Chiv, Jason and Samantha Hambly, Sharon and Gunnar Nielsen, Christoph Dwertmann, Michelle Diack, and other research fellows I met at CSU and NICTA, thank you for the friendship and fun times.
I would like to thank the staff at the School of Computing and Mathematics for making me feel at home during my candidature there. Special thanks go to Sue Kendall, Kerrie Cullis, Chelsea Williams and Jodie Mitchelle, at the School of Computing and Mathematics. To lecturers and friends I met during my master degree at CSU, Geoff Fellows and Ken Eustace, I thank you for your help, inspiration and expertise in technical areas.

Thank you to the research fellows at CSU’s Shut Up & Write, Writing Bootcamp, and Bruce’s Writing Circle, Jasmine MacDonald, Linda Ovington, Soumi Mukhopadhyay, Brook Scriven, and Claire Moloney, who have provided inspiration, friendship, and feedback on my writing. Special thanks go to Cassily Charles, Lisa McLean, and Bruce Gater for organising these studious groups.

To the cyclists in Bicycle Wagga Wagga club, the swimmers at Lunchtime Lappers at the Oasis swimming pool, and the runners at the Wagga Wagga Road Runners, thank you for the fitness and fun.

Extra-special thanks to my parents, Sambo Sar and Sokhom Heng, for their unconditional love.

Thank you to Michelle Hey for proofreading my thesis.

This research has been made possible through the Charles Sturt University International Tuition Payment (CSUITP) over four years, and the Charles Sturt University Postgraduate Research Scholarship (CSUPRS) over three-and-a-half years. My sincere thanks go to all who were involved in helping me to secure funding.
Contents

1 Introduction to the research 2
  1.1 You are what you do online ............................................. 2
  1.2 Need for research .......................................................... 4
  1.3 Research aim and questions ............................................. 5
  1.4 Overview of research designs .......................................... 6
  1.5 Research scope ............................................................ 7
  1.6 Significance of research ................................................ 8
  1.7 Definition of selected terms ............................................ 10
  1.8 Overview of the chapters .............................................. 11

2 Cambodia 12
  2.1 Introduction to Cambodia .............................................. 12
    2.1.1 Facts about Cambodia ............................................. 12
    2.1.2 Politics ............................................................... 13
    2.1.3 Economy .............................................................. 14
    2.1.4 Telecommunications ............................................... 15
  2.2 Culture and tradition .................................................. 15
    2.2.1 Religion ............................................................... 15
    2.2.2 Customs and tradition ............................................ 15
    2.2.3 Life cycle: birth, marriage and death .......................... 16
    2.2.4 Family and gender roles ....................................... 17
    2.2.5 Housing .............................................................. 18
    2.2.6 Cuisine .............................................................. 18
  2.3 Summary ................................................................. 18

3 Information sharing online: the literature 20
  3.1 Introduction .............................................................. 20
  3.2 Intentionally shared information ..................................... 21
    3.2.1 An overview ...................................................... 21
3.2.2 The use of users’ intentionally shared information ......... 22
3.3 Unintentionally shared information .......................... 24
  3.3.1 An overview ........................................ 24
  3.3.2 The collection of users’ unintentionally shared information ...... 25
  3.3.3 The use of users’ unintentionally shared information .......... 27
3.4 Unintentionally shared information and related works ............ 28
3.5 Need for this stage of research ................................ 30
3.6 Research intent and design .................................... 31

4 Users’ unintentionally shared information in the HTTP headers: an exploratory study 32
  4.1 Online activities, sites, and searches used in the study ............... 32
  4.2 Data collection and analysis .................................. 34
  4.3 Findings ...................................................... 36
    4.3.1 First level traverse: From first party sites to third party sites .... 37
    4.3.2 Second level traverse: From third party sites to other third party sites . 39
  4.4 Discussion .................................................... 40
    4.4.1 Tracking by third party sites which are advertisers or data aggregators . 42
    4.4.2 Tracking by third party sites which are SNSs ........................ 42
  4.5 Conclusion .................................................... 43

5 Information sharing in the HTTP headers: 20 participants 46
  5.1 Study intent and design ........................................ 46
  5.2 Cambodian participants ......................................... 47
  5.3 Sampling ....................................................... 48
  5.4 Ethics consideration ............................................ 49
  5.5 Recruitment ..................................................... 50
  5.6 Data collection and analysis .................................... 50
    5.6.1 Wireshark workshop ..................................... 50
    5.6.2 Data collection .......................................... 51
    5.6.3 Data analysis ............................................ 52
  5.7 Findings ....................................................... 53
    5.7.1 First level traverse: from first party sites to third party sites ....... 54
    5.7.2 Second level traverse: from third party sites to other third party sites . 57
    5.7.3 Browser toolbar connections ................................ 59
  5.8 Discussion ..................................................... 61
    5.8.1 Tracking by third party sites which are advertisers or data aggregators . 61
    5.8.2 Tracking by third party sites which are SNSs ...................... 62
5.9 Conclusion .......................................................... 64

6 Social network sites and privacy: the literature 66
  6.1 Social network sites ............................................. 66
  6.2 Social network sites and online users ...................... 69
    6.2.1 The use of SNSs ........................................... 69
    6.2.2 SNS users and the disclosure of information .......... 70
  6.3 Privacy .......................................................... 71
    6.3.1 Definition ................................................ 71
    6.3.2 Personal or private information ......................... 72
    6.3.3 Importance of privacy ................................... 73
  6.4 SNS users and privacy ......................................... 74
    6.4.1 SNSs’ privacy settings and privacy policies .......... 74
    6.4.2 Online tracking and user awareness ..................... 75

7 Social network sites, privacy and research participants: study design 77
  7.1 Intent of research .............................................. 77
  7.2 Methodology .................................................... 78
    7.2.1 Research paradigm ....................................... 78
    7.2.2 Ethnographic method .................................... 80
    7.2.3 Interview technique ..................................... 81
    7.2.4 Sampling and recruiting ................................. 83
  7.3 Data collection ................................................ 83
    7.3.1 Conducting the interview ............................... 83
    7.3.2 Analysis of the interview data ......................... 85

8 Social network sites, privacy and research participants: the findings 86
  8.1 SNS usage ...................................................... 86
    8.1.1 SNS(s) in use ............................................ 86
    8.1.2 Motivation for creating SNS accounts ................. 87
    8.1.3 Browsing activities and technologies .................. 87
  8.2 Intentionally shared information ........................... 88
    8.2.1 Information shared on SNSs and motivation for sharing .. 88
    8.2.2 SNS privacy settings ................................... 89
    8.2.3 Privacy policies ......................................... 90
  8.3 Privacy ........................................................ 91
    8.3.1 Definition of privacy .................................... 91
    8.3.2 Personal or private information ......................... 91
11.2 Limitations of research ............................................ 138
11.3 Future research ......................................................... 140

References ........................................................................... 142

A Human Research Ethics Committee Approval .................. 156
B Information sheet ............................................................. 159
C Consent form ................................................................. 162
D Interview Questions ......................................................... 164
E List of participants ......................................................... 167
Certificate of authorship

I hereby declare that this submission is my own work and to the best of my knowledge and belief, understand that it contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at Charles Sturt University or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by colleagues with whom I have worked at Charles Sturt University or elsewhere during my candidature is fully acknowledged.

I agree that this thesis be accessible for the purpose of study and research in accordance with normal conditions established by the Executive Director, Library Services, Charles Sturt University or nominee, for the care, loan and reproduction of thesis, subject to confidentiality provisions as approved by the University.

- Name: Rath Kanha Sar
- Signature:
- Date:
List of Figures

2.1 Cambodia map ........................................ 13
3.1 Trove Facebook third party application ..................... 23
4.1 File captured by Wireshark .................................. 34
4.2 Ebay traverse map ......................................... 40
4.3 The traverse of the browser’s connection ...................... 41
5.1 narutoget.com traverse map .................................... 58
6.1 Timeline of the launch dates of major SNSs .................... 68
9.1 Managing Facebook’s widget ................................... 118
10.1 The flow of an online user’s unintentionally shared information .......................... 127
List of Tables

3.1 HTTP Conversation between the browser and Google ................................................ 29
4.1 Online activities and sites chosen for the studies ......................................................... 33
4.2 First party site connects to third party site while also sharing user’s search terms ........... 36
4.3 List of third party sites connected to by the browser while visiting first party sites .......... 37
4.4 List of third party sites connected to by the SNS ......................................................... 38
4.5 The leakage of the researcher’s LinkedIn ID and full name ....................................... 38
4.6 Lyrics007 shares the researcher’s search keyword to Twitter while retrieving Twitter widget ................................................................. 39
4.7 The traverse of ebay the researcher’s search keyword ................................................. 39
4.8 Lyrics007 traverse ...................................................................................................... 40
4.9 Summary of the information leakages in the study ...................................................... 41
5.1 Participants’ online activities and sites ........................................................................ 52
5.2 Seila’s browsing at narutoget.com  ............................................................................. 53
5.3 List of third party sites ................................................................................................ 54
5.4 News Web sites sharing to third party sites ................................................................. 55
5.5 gsmarena.com HTTP headers ..................................................................................... 55
5.6 Participants’ Facebook connections ........................................................................... 55
5.7 Phnompenhpost shared Phara’s visited page to Google ............................................... 56
5.8 Visited sites and SNSs (F = Facebook, T = Twitter, G = Google) ................................. 56
5.9 HTTP headers: Seila’s browsing propagation ............................................................... 57
5.10 Seila’s browsing propagation ..................................................................................... 58
5.11 Makara’s information propagation ............................................................................. 59
5.12 Phanna’s information propagation ............................................................................ 59
5.13 Nary’s toolbar connection establishment .................................................................. 60
5.14 Babylon connection .................................................................................................. 61
5.15 Incredibar’s HTTP connection to third party sites ..................................................... 61
5.16 Participants’ toolbars and third party sites ............................................................... 61
5.17 Moly’s browsing activities and associated third party sites ..................................... 62
5.18  Moly’s visited pages able to be tracked by Scorecard Research ............. 62
5.19  Kosal’s browsing activities and associated third party sites .................. 63
5.20  Kosal’s browsing activities able to be tracked by SNSs ....................... 63

E.1  List of participants ................................................................. 168
Chapter 1

Introduction to the research

1.1 You are what you do online

This research begins with an exploratory study (detailed in Chapter 4), with one of the outcomes indicating that online users’ activities can be tracked and recorded by numerous sites including advertisers, data aggregators, and popular social network sites (SNSs) such as Facebook, Twitter, and Google Plus. Online tracking helps advertisers to provide tailored advertisements on the sites visited by users. However, tracking by SNSs appears to be worrisome because SNSs are now able to collect and store not only personal information voluntarily provided by users, but also information about users’ activities outside SNS boundaries. It is not clear how information will be used once collected by SNSs. The combination of collected information may reveal too much about a person’s life. Let us consider the scenario of a female Cambodian named Soriya.

Soriya is very nice and helpful. She has just finished her undergraduate degree and is now looking for a job. She is updating her profile on LinkedIn, an SNS dedicated to professionals. She is also using Facebook to keep in touch with her friends and relatives. Her Facebook activities include exchanging messages with her Facebook friends, and posting pictures, comments and status updates about her life. In addition, she is also playing interactive third party applications such as FarmVille and Mafia wars (by Zynga1) on Facebook. Sadly she has a friend who is HIV positive and depressed, and is having problem with drugs and alcohol. In order to provide emotional support to her friend, Soriya spends some time searching online for information to help her understand her friend’s condition. Meanwhile, Soriya’s cousin is pregnant; hence Soriya also searches online for useful information related to pregnancy such

1http://zynga.com
as diet and exercise. She also orders some books and vitamin supplements from online stores for her pregnant cousin.

As suggested by the findings within this research, Soriya’s online activities are being tracked and recorded by unknown advertisers, data aggregators and SNSs like Facebook. Thanks to this tracking, Facebook is able to display the right ads on Soriya’s profile based on her browsing activities and she is offered books and vitamins at discounted prices from online stores. In addition, thanks to Facebook default privacy settings which make users’ profile pages publicly available and searchable, Soriya’s childhood friends are able to find and reconnect with her based on her basic personal information including pictures, her hometown and high school. The Facebook platform also allows Zynga to provide interactive and social online games such as Mafia Wars and FarmVille. Soriya and other SNS users enjoy playing these types of games with their friends on Facebook, in exchange for their information being collected or harvested by Zynga and other unknown parties.

The combination of Soriya’s information, which includes her SNS profile information, her interaction with her friends as well as her online activities, reveals so much about her life. This information could be embarrassing or damaging to her future job opportunities in Cambodia. For example, her job or health insurance application may be rejected on the assumption that she may be pregnant, or involved with HIV, drugs, and alcohol. Of course not all the collected information is an accurate representation of Soriya’s life. She is neither pregnant, nor HIV-positive, but the aggregated data from her online browsing can not distinguish for whom she did the searches.

In order to determine how privacy concerns are raised by Web sites visited by Soriya, Soriya’s story can be analysed based on two ethical theories: utilitarianism and deontology. From a utilitarianism viewpoint, the expected outcome or consequences of an act is very important to determine whether or not that act is morally acceptable, while the role of duty and respect for persons are the key to what is morally permissible for a deontologist (Tavani, 2011). Two outcomes can be drawn from Soriya’s story.

1. The act of sharing and gathering users’ information among visited sites (e.g., Facebook and LinkedIn) in order to enhance their online experience (e.g., friendship reconnection, relevant ads, and interactive games) appears to be consistent with both Kant’s principles of treating people as ends in themselves and the principles of utilitarianism which promote the greatest good for the greatest number.

2. However, from a deontological viewpoint, users are not being respected as persons and are being treated as a means to end (the means being users’ information while the end is the income generated from the sale of that information). Their informational privacy
is violated given their inability to control who is collecting the information, the type of information being collected about them, and for what purpose the information will be used. This can also result in a negative online image or reputation, as well as being damaging to career opportunities for online users - which is not morally permissible according to utilitarianism.

Analysing Soriya’s story using ethical theories reveals both positive and negative impacts brought by the sharing and tracking of online users’ information.

1.2 Need for research

SNSs usually encourage users to disclose and share different types of information while staying connected with other SNS users or friends. Currently, Facebook is by far the most popular SNS, attracting more than one billion active users by mid 2013 (Statistic Brain, 2013). As briefly discussed in Section 1.1, the tracking activities of SNSs appear to be problematic as they are holding a large volume of information that may reveal much about a person’s life (as in Soriya’s scenario, with her personal information, her activities on Facebook, and her browsing activities among other sites being tracked).

A number of studies shed light on how users managed their privacy and information disclosure online (Livingstone, 2008; Young, 2009; Krasnova, Spiekermann, Koroleva, & Hildebrand, 2010; Lawler & Molluzzo, 2010; Al-Saggaf, 2011). Other studies surveyed users’ perceptions and attitudes towards online tracking (McDonald & Cranor, 2009, 2010a; Wills & Zeljkovic, 2011). Other studies also shed light on the meaning and perception of privacy (Yao-Huai, 2005; Kitiyadisai, 2005; Tamura, 2005; Ess, 2005). These studies provide an understanding of the online behaviours of users from different backgrounds and cultures, including American, Australian and Middle Eastern, and how culture, gender, and age affect users’ disclosure of information online and their tolerance towards online tracking. Existing works have not yet studied the perspective and awareness among users from developing countries like Cambodia where the Internet is not widely accessible. Despite this, SNSs are popular among Internet users in that country.

Existing technical literature focused on the leaking or sharing of users’ information from sites visited by users to other sites (Krishnamurthy & Wills, 2006b; Krishnamurthy, Malandrino, & Wills, 2007; Krishnamurthy & Wills, 2009, 2010a; Mayer, 2011). This leaking or sharing enabled other sites (that are not visited by users) to track users’ online activities. However, the results of those studies did not closely reflect the browsing habits of users in real life. They employed a large number of sites when users are not likely to browse a large number of the
same category of sites simultaneously, such as 10 SNSs at a time, but rather a combination of different categories instead, such as an SNS, online shops, e-mail, and a search engine (Purcell, 2011). Some users can be more privacy savvy than others; hence, they have their browsing in a private mode (Bursztein, 2012). To the knowledge of the researcher, no previous study has taken users’ browsing habits into consideration.

Meanwhile, the impact of technologies has been analysed by employing philosophical theories or frameworks to elicit the ethical implications. For example, the Contextual Integrity (CI) framework has been extensively used in the literature to analyse online privacy in several contexts including public records online and radio frequency identification (RFID) (Nissenbaum, 2004), data mining (Nissenbaum, 2004; Tavani, 2007b), blogosphere (Grodzinsky & Tavani, 2010), cloud computing (Grodzinsky & Tavani, 2011) and Facebook’s News Feed feature (Hull, Lipford, & Latulipe, 2011). The CI framework (detailed in Chapter 9) provides a set of guidelines that can be used to evaluate the impact of a technology or system from a moral viewpoint. To the knowledge of the researcher, the CI framework has not yet been used in analysing the practice of tracking users’ online activities by SNSs.

Based on the rationale above, there is a need to bring attention to the impact of the practice of tracking users’ online activities by SNSs, particularly among users from developing countries like Cambodia where Internet use has not previously been researched. Although Cambodia is a poor country, the Internet is accessible by the younger generation such as university students and SNSs are widely used by Cambodian internet users. Additionally, there is also a need to investigate the online information sharing resulting from users’ actual browsing activities, while at the same time gauging their perception and reaction to the practice of SNS tracking of their online activities. Online users are affected by tracking and affected individuals deserve ethical consideration (Australian Law Reform Commission, n.d.). As briefly discussed in Section 1.1, philosophical theories help to identify both positive and negative impacts brought by a technology or a practice. Therefore, this research also employs the CI framework in order to expand the analysis of the practice of SNS tracking, in terms of ethical considerations.

1.3 Research aim and questions

This research aims to explore the impacts of information sharing via the HTTP headers of Web sites visited by users from a technical, social and philosophical perspective. It is important to include these three perspectives because, as considered in Section 1.2, there is a need to find out what information about users is being shared, to what extent users are aware of this

---

2http://www.alexa.com/topsites/countries/KH
sharing and how they feel about it, and what ethical implications are raised by such a practice. Firstly, the ramifications of information sharing in HTTP headers will be investigated in an exploratory study (Chapter 4) before applying the same method to research participants in Cambodia (Chapter 5). Secondly, participants’ perception and awareness of online privacy and online tracking will be ascertained, while last but not least, this research will employ the CI framework to analyse the action of online tracking from an ethical perspective.

As noted earlier, the researcher conducted an exploratory study to analyse the HTTP messages resulting from online browsing activities, the outcome revealing that online activities can be tracked and recorded by various agents or Web sites such as advertisers, data aggregators and particularly, SNSs (Facebook, Twitter, and Google Plus). Based on this result, the researcher narrowed the focus of the research into the case of tracking users’ information by SNSs. Therefore, this research aims to address the following research questions:

1. What is the nature of information sharing in HTTP headers?
   (a) What type of information is being shared?
   (b) With whom is this information being shared?
   (c) What are the implications of this sharing?
2. What are SNS users’ views on online tracking and privacy?
   (a) What does privacy mean from SNS users’ perspectives?
   (b) To what extent are they aware of online tracking in general as well as online tracking by SNSs?
   (c) What are users’ reactions to the practice of sharing and tracking information online?
3. What are the ethical implications of collecting and tracking users’ information by SNSs?

1.4 Overview of research designs

As mentioned previously, this research began with an exploratory study which allowed the researcher to not only assess the feasibility of conducting the experiments among research participants, but also to shape the research scope and research questions for this thesis. The exploratory study involved analysing HTTP messages produced from the researcher’s online browsing on different Operating Systems (OSes) such as Windows, Mac OSx, and Linux, while following a set of browsing protocols and activities. These browsing activities took place between November 2011 and January 2012. This study is necessary to assess the feasibility of the experiments as well as the associated risks; for example, the risk of passwords being revealed in the recorded files.

Once the possible risks in the experiments were assessed, and when ethics approval from
Charles Sturt University was obtained, the data collection was conducted in Cambodia with 20 Cambodian participants. The first stage of this research involved experiments where participants performed online browsing activities while having the HTTP traffic recorded. This stage took place on the 28th of July 2012. The recorded files were analysed by the researcher in order to identify the types of information being shared or leaked in the HTTP headers to third party sites from sites visited by participants.

The second stage of this thesis employed a qualitative approach to research. It was influenced by ethnographic methods as it studied participants’ thoughts and opinions. Open-ended structured interviews were conducted with 20 participants where they were told about the outcomes of their online browsing experiments. All of them were asked the same series of questions and their answers were analysed using NVivo Version 9. The interviews took place between the 6th of August 2012 and the 11th of August 2012.

The last stage of the research involved analysing the practice of online tracking of SNSs such as Facebook, Twitter and Google Plus from a philosophical viewpoint by applying Helen Nissenbaum’s CI as a framework. The analysis employed the framework’s norms as well as its decision heuristic in order to determine whether or not the practice of tracking online users’ activities by SNSs violates their right to privacy.

1.5 Research scope

This research focused on the collection of online users’ information and the associated impacts created by this practice. A large volume of user information is available online; however, this research did not investigate the impacts of information disclosed online by users among SNSs (termed by the researcher as “intentionally shared information”). Rather, it investigated users’ information that is involuntarily given or disclosed by them, but can be available to other parties that are unknown to users. The researcher termed this type of information as “unintentionally shared information” (more information on intentionally and intentionally shared information will be presented in Chapter 3). This study observed only the information sharing in the HTTP headers resulting from participants’ online browsing, but it did not observe information sharing via other means such as flash cookies, or API (Application Programming Interface) that enables interaction and information sharing among SNSs and third party applications.

Data collected in the experiments was from participants’ personal computers. The researcher did not collect the HTTP data from other devices such as mobile phones, or tablets. Hence, online tracking uncovered in the experiments was limited to the browsing via personal computers among browsers such as Firefox, Chrome, and Internet Explorer, that were used by research
participants. Online tracking in this research refers to tracking by Web sites, not by other internet users who may track and stalk other users online. Hence, it did not examine the criminal issues such as online stalking, or identity theft. Some of data collected from the HTTP headers were encrypted; however, this research investigated only unencrypted data.

The findings of the experiments in the exploratory study as well as from research participants revealed that information about users’ online activities is collected by numerous sites including advertisers, data aggregators and SNSs. Participants’ browser toolbars were also observed to transmit users’ information to third party sites. In addition, users’ information was noted to be dispersed from one third party site to other third party sites. This research did not analyse the tracking by advertisers or data aggregators. It did not focus on first party sites or browser toolbars that were also stakeholders in the tracking. It focused only on the tracking by SNSs. Hence, participants’ interviews as well as the ethical analysis centred around the issues of tracking by Facebook, Twitter and Google Plus.

This research is influenced by ethnographic method, and aims to elicit participants’ opinions regarding their awareness and perception of online tracking and privacy. Hence, ethnography method, whereby users’ opinions can be collected by different techniques, was briefly mentioned in this thesis. Only open-ended structured interviews were used with participants in Cambodia. In addition, the answers captured in the interviews were derived from the outcome of experiments that were based on participants’ browsing activities.

In addition, although other theories like Restricted Access/ Limited Control (RALC) and ontological interpretation of informational privacy were introduced in this research (Chapter 9), this thesis employed only the CI framework introduced by Helen Nissenbaum. Hence, the analysis of the tracking by SNSs in this research was based on the guidelines and principles of the CI framework. In addition, this thesis focused only on the ethical implications of online users’ privacy; hence, legal consideration or justification is not within the scope of this research.

### 1.6 Significance of research

SNSs have become increasingly popular and have attracted more than one billion active users worldwide to date. Issues and concerns about information disclosures on SNSs have been discussed in the existing literature as well as in the news. For example, star of the *Harry Potter* movie, Daniel Radcliffe, recently mentioned that he avoided SNSs for privacy purposes as this helped him spark less interest, particularly in the media (Child, 2013). Meanwhile, online tracking has been in existence ever since HTTP cookies were invented in 1994. However, the tracking of online users’ behaviours or the collecting of online users’ unintentionally shared
information, as discovered in this research, is relatively new. Hence, it is less often reported on in the media compared with the issues of information leakage from SNS users to other SNS users.

This research is the first that both allowed Cambodian research participants to engage in browsing experiments and observe their information sharing in HTTP headers, as well as gauge their perception and awareness of the practice of online tracking. Unlike existing literature, information sharing detected in this research is based on real life users online browsing activities. The interviews revealed that Cambodian research participants were not aware of online tracking by third party sites. However, they cared about and valued their privacy; and expressed concern about the loss of privacy that resulted from such a practice.

Additionally, this is the first research that involved analysing the impacts of tracking by SNSs from both the perspective of online users, as well as an ethical perspective by employing CI as a framework. The framework suggested that tracking and recording users’ information without their informed consent violates their right to privacy. The findings from both perspectives are consistent. Hence, tracking of users unintentionally shared information by SNSs appears to be wrong.

Analysing the current practice of tracking from the point of view of online users as well as from the moral perspective allowed a deeper understanding about the impacts of the practice on stakeholders including users and society as a whole. This research provided insight into participants’ feelings and reactions toward the practice of online tracking. The guidance offered by CI also gave an appreciation of the privacy expectations within the context of SNSs and the tracking of users. Hence, it may aid future work if proper protection for the welfare, rights, dignity and safety of online users is afforded.

The researcher envisages that this research will pave the way for future research that adds to the knowledge base about online information sharing. In addition, the outcomes of this research demonstrated that SNS users do care about their privacy and were not happy about being tracked. As online users are SNS customers (given that SNSs’ business models rely heavily on users’ information), it is important that SNSs keep their customers happy. Notwithstanding, future work may also assess if SNSs remain popular once users are aware of tracking. The Internet has a positive influence on our lives; therefore, it is crucial to detect and identify the risks and mitigate them, in order to maximise the potential benefits brought by this technology. Last but not least, the researcher sincerely hopes that this research will add to the current literature about Cambodia in terms of Internet usage, and bring Cambodia to the attention of other researchers, particularly those who are Cambodian.
1.7 Definition of selected terms

- **CDN** (Content Delivery Network or Content Distribution Network): refers to a large interconnected system of computers or servers that rapidly provides Web content to users based on proximity, as this system is deployed in multiple data centres across the network.

- **Clickstream**: a series of mouse clicks by a user during a visit to a site. A user’s clickstream can be recorded or accumulated by the site owner or third party sites for marketing purposes.

- **Contextual Integrity (CI)**: introduced by Helen Nissenbaum, this is a privacy framework or model that helps to understand the expectations of privacy by providing guidance on how to respond to conflicts between values and interests within a context where moral indignation arises.

- **First party site**: for the purpose of this research, first party sites refer to any Web site that is directly or intentionally visited by online users.

- **HTTP**: HyperText Transfer Protocol is an application-level protocol that provides a set of standards for Web browsers and Web servers to communicate via HTTP request and response messages. It is a stateless protocol, so it uses HTTP cookies to remember the state.

- **HTTP Headers**: are core components of the message header of requests and responses in HTTP protocol. They carry information about the browser, the requested page, and the server (an example can be found in Section 3.4).

- **HTTP cookie**: a small piece of data sent from a Web server and stored on a user’s device (a computer, mobile phone or tablet) while user is visiting that Web site. An HTTP cookie is designed to remember the state, for example, to notify the Web server that the same user has returned, because HTTP is a stateless protocol.

- **ICT (Information and Communications technology)**: is often used as an extended synonym for information technology. It is concerned with the uses of digital technology (e.g., computers and telephones) to store, retrieve, transmit, or manipulate the digital data.

- **Social network sites (SNSs)**: “Web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system” (boyd & Ellison, 2007, p. 122).
Third party site: for the purpose of this study, a third party site refers to the site that is not directly or intentionally requested by users. For example, third party sites can be advertisers which populate advertising contents on the first party sites visited by users.

1.8 Overview of the chapters

The thesis is organised as follows. Chapter 2 provides an overview of Cambodia and its culture. Chapter 3 presents background and literature about information sharing online. It also advances justification for the study based on the identified literature gap, and it outlines the study intents and designs. Chapter 4 details the steps and processes in conducting the exploratory study and presents its outcomes. Meanwhile, Chapter 5 details the steps and processes in undertaking online browsing experiments with research participants in Cambodia. It includes a brief introduction to research participants, ethics consideration, participant recruitment, and outcomes of the experiments. Chapter 6 presents literature and background related to SNS and privacy, while Chapter 7 provides the method and techniques used in conducting qualitative interviews with Cambodian participants. The findings of the interviews are presented in Chapter 8. The practice of tracking by SNSs is analysed from the perspective of the CI framework in Chapter 9. Chapter 10 recapitulates the research questions and discusses the findings within this thesis, while Chapter 11 concludes the thesis as well as outlining possible future works which could be derived from the findings of this research.
Chapter 2

Cambodia

As this research focuses on participants from Cambodia, this chapter provides an overview of the country and its culture. It starts with an introduction to the country’s history, politics, and economy in Section 2.1. Section 2.2 provides an overview of the country’s culture and tradition, which includes its religion, customs, lifestyle, and cuisine.

2.1 Introduction to Cambodia

2.1.1 Facts about Cambodia

Cambodia, known as the Kingdom of Cambodia, is a country in South East Asia with a population of about 15 million in 2013 and a total landmass of 181,035 square kilometres (CIA, 2013). Like most other South East Asian countries, the tropical climate in Cambodia is humid and warm almost all year round with an average temperature range from 21 to 35 degrees Celsius. There are two seasons known as rainy or monsoon reason (between May and November) and dry season (between December and April) (MOT, 2013).
The country is a constitutional monarchy with Norodom Sihamoni as the head of state, and its Prime Minister, Hun Sen, as the head of the government. Cambodia shares borders with Vietnam to its east, Laos to its northeast, Thailand to its west, and the Gulf of Thailand to its southeast, as shown in Figure 2.1 (CIA, 2013). The geography is dominated by the Mekong River and the Tonle Sap, the main source of fish. Phnom Penh is the capital and largest city of Cambodia. The population of Cambodia consider themselves to be Khmer, descendants of the Angkor Empire. Khmer is also the country’s official language. The language is derived from the Mon-Khmer (Austro-Asiatic) language family (MOT, 2013). Foreign languages such as French, English and Mandarin are also spoken among the Cambodians. French is widely spoken and understood among Cambodian elderly, while English is more popular with the younger generation and Mandarin is used by the Chinese descendants in Cambodia.

![Figure 2.1: Cambodia map](image)

2.1.2 Politics

Cambodia’s political underpinnings have been unstable and dominated by unpredictable leadership and policies. The Angkor Empire was the most powerful empire in South East Asia and
extended over much of the area between the 10th and 13th centuries, but experienced a down-
fall as a result of the attacks by the Thai and Cham (currently Vietnam) (CIA, 2013). Cambodia 
was a French colony between 1863 and 1953, and became part of French Indochina in 1887. 
Norodom Sihanouk ruled the country as an autocracy until 1970 when General Lon Nol estab-
lished a military government and removed Sihanouk from power (Wetzel, 1995).

In April 1975, the country experienced another struggle under the Communist Khmer Rouge, 
led by Pol Pot, who reacted to the military rule of Lon Nol (Wetzel, 1995; CIA, 2013). The 
Khmer Rouge (or Red or Communist Cambodians) abolished both the Lon Nol government and 
the monarchy, and evacuated all cities and towns. At least 1.5 million Cambodians died either 
from execution, brutal labour camps, disease or starvation between 1975 and 1979. The first to 
be killed were Buddhist monks, urban inhabitants, government officials and people with tertiary 
or Western education (e.g., academic staff or those who graduated from overseas educational 
institutions).

In 1978, Vietnam invaded and occupied the country, and drove the Khmer Rouge into the 
countryside. The 1991 Paris Peace Accords mandated democratic elections, sponsored by the 
United Nations, held in 1993. The nation was reunited again under the monarchy in 1993, with 
Norodom Sihanouk as the head of state. The remaining Khmer Rouge surrendered in 1999 and 
are awaiting trial for crimes against humanity. King Norodom Sihanouk renounced his throne 
in 2004 and selected his son, Prince Norodom Sihamoni to succeed him. In October 2012, 
Former King Norodom Sihanouk passed away after a heart attack in Beijing where he had been 
receiving medical treatments (ABCNews, 2012).

2.1.3 Economy

Despite the unstable political environment and civil war, Cambodia experienced a period of ro-
bust economic growth of 6% between 1993 and 2003, and 11.1% between 2004 and 2007 (UN, 
n,d.). Since 2004, the country’s economy has relied heavily on garment exports, agriculture, 
tourism and the construction industry (CIA, 2013; UN, n,d.). Economic growth has reduced 
the national poverty rate from 47% to 30% in 2007 (UN, n,d.).

Nevertheless, Cambodia still remains one of the poorest countries in Asia with one-third of the 
population still living below the poverty line (less than US$1 per day). Of this, 12% are reported 
to be facing hunger and food insecurity, while 37% of Cambodian children under the age of 5 
are suffering from malnutrition (CIA, 2013). Economic development is said to be inhibited by 
corruption, limited educational opportunities, poor job prospects and income inequality.
2.1.4 Telecommunications

Landlines and cellular services are adequately accessible in Cambodia (CIA, 2013). The country continues to grow, and so does its connection to the rest of the world. Mobile phone coverage and Internet accessibility is rapidly expanding all over the country. In 2011, approximately 13 million Cambodians owned a mobile phone. Internet access is available in many locations such as coffee shops, bars, restaurants, and Internet cafes. Improvements to Internet connections and competition between Internet Service Providers (ISPs) have resulted in lower prices, for example, from only U.S$12 a month plus the price of modem rental (PPCTV, 2013). However, although only 3.1% of the population (approximately about 78,500 people) are reported to have access to the Internet (WB, 2011), an estimated 700,000 people use Facebook (CambodiaWC, 2012; CH, 2013; Khoun, 2013) and the majority of users are under 35 years of age (CH, 2013).

2.2 Culture and tradition

2.2.1 Religion

Cambodian culture and tradition have a rich history dating back many centuries. Throughout the history of Cambodia, religion has been a major source of cultural inspiration. It is believed that the Khmer people developed a set of unique traditions from the syncretism of indigenous Buddhism and Hinduism (MOT, 2013). The country’s official religion is Theravada Buddhism which is practiced by approximately 90% of the population. Muslim Cham, Christianity and small hill tribes represent minority religions in the country (MOT, 2013). Cambodian people are spiritual and they believe in the unseen world and the power of the spirit. Regardless of religion or belief, Cambodia maintains a harmonised state. Buddhist monks play an important role in Khmer society (FRD, 1987). They are called upon to perform a number of functions at festivals and other ceremonies related to birth (e.g., infant naming ceremonies, or baby blessings), marriage, death, and at other celebrations including house warmings, and store openings.

2.2.2 Customs and tradition

Etiquette and protocols are used in Khmer society to maintain a sense of common harmony. Khmer culture is very hierarchical. Khmer people are taught to pay respect to elders. This is usually reflected in the language that has different styles of usage across three broad categories including royalty, clergy and laity (FRD, 1987). For example, the word keng or dek (to sleep)
is used in general among the laity, while *sam-ran* is used when speaking of elders, and *pa-tum* is used when speaking of royalty.

Traditionally, Khmer people greet and show respect to one another with a *Sampeah* which involves placing the palms together in front of the chest or face while slightly bowing with a polite *Chumreap Sour*. There are also different ways to *Sampeah* depending on age, relationship and hierarchy (Kwintessential, n.d.; Keo, 2008b; MOT, 2013). For example, the higher the hands are held, the more respect is conveyed. The practice of shaking hands is becoming more widely adopted in Cambodian society, particularly with foreigners, among men, or among government officials. However, *Sampeah* is still practiced and is more appropriate, especially between different genders or between foreigners and Khmer women.

### 2.2.3 Life cycle: birth, marriage and death

The birth of a child is a happy event for a family. Pregnant women, particularly in the rural areas, are expected to respect a number of taboos associated with restricted foods during pregnancy and immediately after birth, for the benefit of both the baby and the mother (FRD, 1987; Montesanti, 2011). For example, *hot* or *spicy* food is considered dangerous and should not be consumed during pregnancy and breastfeeding. The ceremonies of the dead and homes of sick people should also be avoided by pregnant women. The death of pregnant women or women in labour is considered to be evil - these women are believed to become evil spirits.

The choice of spouse in Cambodia is a complex process. Marriage and courtship processes usually involve both parties’ parents, friends, relatives, and possibly a *matchmaker*. The courtship ritual is different in the city and the country in terms of the attitudes to courtship and the choice of spouse. Larger cities are seen to be more influenced by western societies than in the rural areas and some parents are happy with their children choosing their own spouses or dating people they like before taking any further step into *marriage*. However, *premarital sex* is taboo (FRD, 1987) in Khmer culture. Nevertheless, arranged marriages still exist in Cambodia, both in rural and urban areas (Keo, 2008a).

The Khmer wedding is yet another complex and colourful process. It involves many steps including ritual hair cutting, tying holy cotton threads around the bride’s and groom’s wrists, and passing a U-shaped candle around a circle of happily married couples to bless the newly wedded couple. Western suits, bridal dresses and wedding cakes are also very common, especially at an evening reception. Khmer weddings may also include the Chinese traditional *tea ceremony* if the bride’s and groom’s families are of Chinese descent. The wedding would last for three days in earlier times, but will now only last one-and-a-half days particularly in the
cities, due to the demands of modern living (FRD, 1987).

*Black* and *white* are mourning colours. The spouse and children of the dead show mourning and respect by shaving their heads and wearing all white. Buddhist Khmer are cremated, while some are buried. In Khmer society, death is viewed as the end of one life and the beginning of another, hopefully a better one (FRD, 1987).

### 2.2.4 Family and gender roles

The roles of Cambodian men and women are given a great deal of respect in traditional Cambodian culture. Legally, the husband or father is the head of the family, while to some extent, the wife has authority in family economic matters. Now mostly in rural areas, the husband is the breadwinner and is responsible for providing shelter and food. Meanwhile, the wife is mainly in charge of household chores such as cooking, cleaning, and taking care of the children. In the case of farmers, both husband and wife work in the rice field or on the farms (FRD, 1987). However, when society changes, roles will also change. In modern society, both husband and wife go to work, and they share family responsibilities.

Traditionally, Cambodian females are taught at a very young age to be polite, obedient, gentle and softly spoken, and to avoid breaking any moral principles (Keo, 2008b). In the very old times, when reaching puberty, girls were expected to participate in a ritual called *Chol Mlub* (direct translation: entering the shade) (FRD, 1987). *Chol Mlub* is like an internship of approximately six months where females were trained to be mentally mature women by their mothers. They were isolated from the rest of the family in a separate building or room where men were not allowed to enter. They learned to do housework (e.g., cooking) and to take care of themselves (e.g., hair and skin). After the period of seclusion, they were considered marriageable.

Cambodian children have an obligation to their parents to do their utmost to maintain their parents’ honour. Returning gratitude to their parents is a social and cultural expectation. At school, Khmer children are taught to be polite and obedient, and pay respect towards elders and Buddhist monks (FRD, 1987; Keo, 2008b). They are also taught about *Chbap*, or didactic codes where there is specific advice for daily life for different groups such as *chbap pros* for boys, *chbap srey* for girls, and *cheap chav* for grandchildren, to name a few (Ledgerwood, n.d.). From the researcher’s experience, she was expected to remember by heart those *chbap* or codes. For example, females are advised to walk and talk quietly. Both genders are advised not to talk with others about family matters, as it may bring embarrassment to the family.
2.2.5 Housing

Both nuclear and extended families are equally popular in Cambodia. Nuclear families are more common in the cities. However, a newly wedded couple usually lives with the bride’s family until they can afford to buy a house. In the rural areas, Khmer people live in rectangular houses from six to ten metres square in size, constructed of a wooden frame with walls of bamboo (FRD, 1987). In the city, on the other hand, flats are very common. Better-off families usually own a villa - a big house with multiple rooms and a front and backyard. Most affordable flats in the city consist of one or two bedrooms, a kitchen, a common area, and a bathroom. In the city areas, it is very common for people to live and conduct their business within a flat at the ground floor. It is also common for children of the family to share the communal sleeping space.

2.2.6 Cuisine

Cambodian cuisine is a unique blend of flavours and colours that enhance the natural ingredients. A very authentic Khmer spice paste is known as Kreung which is a blend of native ingredients such as galangal, turmeric, garlic, shallots, lemongrass, and kaffir lime leaves. Kreung can be used as a meat marinade for barbecues, stir-fries or stews. Another ingredient that makes Khmer cuisine unique is Prohok. Prohok is a pungent fermented fish paste that is always present in any authentic Khmer dish. Cambodians also use other fermented ingredients like Kapi (fermented prawn paste) and fish sauce. Collectively, Kreung, Prohok, Kapi, and fish sauce have become an important aromatic combination used in Cambodian cuisines (FRD, 1987; MOT, 2013)

A typical Cambodian meal includes rice, the staple food for Cambodians, with one or a few other dishes. The dish can be a soup (Somlor), stir-fry, grills, or salads. Even though noodles are very popular in Cambodia, a meal is always served with a bowl or plate of rice. Chilli is optional and is added to individual taste. Meanwhile, Cambodian desserts consist of fresh produce like tropical fruits (e.g., melon, palm fruits, mango, jackfruits), sweet potato, pumpkin, and seeds like green, red or black beans, and coconut milk.

2.3 Summary

Despite the fact that Cambodia has enduring significant hardships, the country is striving to rise above the poverty line. Khmer culture and social norms and traditions are taught through
educational institutions as well as families. Khmer people are taught at very young age to be gentle and polite, and to pay great respect to elders. Each family member plays a different role to keep the family harmonised. Cambodian males are the head of the family while Cambodian females takes care of the family finances and household chores. However, societal and family roles are changing. For example, many husbands and wives are now working full time, so it is quite common that they share the household chores.

The Cambodian economy is growing and more people are able to access education and telecommunications. The participants in this research were recruited from Cambodia. It is hoped that by undertaking research focusing on Cambodian participants, the researcher will discover and contribute new knowledge to the existing and scarce literature about Cambodia.
Chapter 3

Information sharing online: the literature

This stage of the research investigates information sharing in HTTP headers resulting from online users’ browsing activities. The researcher first conducted an exploratory study (details in Chapter 4) before conducting larger scale experiments among research participants from Cambodia (details in Chapter 5). To provide context for the exploratory study and large scale experiments, this chapter examines the literature on information sharing online. Section 3.1, Section 3.2, and Section 3.3 provide background on the types of information that are currently available online. Section 3.4 evaluates the existing literature while Section 3.5 discusses the current gaps in the literature and the need for research. Meanwhile, Section 3.6 details the first stage of the research’s intent and design.

3.1 Introduction

The Internet becomes an extremely useful source of information as a large amount of information is added to the online environment every day. It offers a variety of benefits to users due to the fact that many activities can be done online. These include Internet banking, online teaching and learning, online shopping, and sharing of information and resources online, to name a few. The Internet also enables communication through different platforms such as e-mail services, Web forums, instant chat services (such as Yahoo messenger, GTalk, and MSN), as well as SNSs.

In addition to the benefits provided by the Internet, there are concerns about its impact when different types of information about individuals is being rapidly collected, in large volumes,
without users being aware of the collection, how their information will be used, or the duration of time for which information will be kept (Milne & Culnan, 2004; Tavani, 2011). More and more personal data is generated or generously provided by online users, and large volumes of information are available in topical databases or diffused online. Potential sources of online information are very diverse. User information is scattered among many services and applications online. From the perspective of online users, this divulged online information can be classified into two types: “intentionally” and “unintentionally” shared information. These types of information, once diffused or shared online, can be collected and transferred to third party organisations with whom the users do not have any established relationship.

3.2 Intentionally shared information

3.2.1 An overview

The term “intentionally shared information” refers to information about users that is voluntarily supplied by them online to an intended recipient group or with an expectation that this information can be viewed by others. This type of information can include name, pictures, videos, thoughts or opinions, and contact details such as postal address, e-mail address and phone numbers.

Intentionally shared information is observed to be disclosed to some sites by users in order to become a member and to be able to view the site content or updates. Some registrations can be necessary for content that requires payment. For example, although users have free access to Sky FM\(^1\), an online radio station, they need to sign up for a **Premium** account and pay US$4.99 monthly in order to access more than 65 radio channels with higher quality audio and without any advertisements. However, some sites require registration for contents that do not require any payment. For example, Ozsale\(^2\) and Brands Exclusive\(^3\) require that users register with their full name and e-mail address in order to view sale items, without necessarily making any purchase or payment.

Users also intentionally share their information including thoughts and opinions in Web forums, blogs, or SNSs. SNSs are sites where large volumes of personal information are being voluntarily shared by SNS users; for example, there are more than one billion users, one-seventh of the world’s population, who are actively sharing and generating information on Facebook (Brenner & Smith, 2013). As defined earlier in Section 1.7, SNSs can be defined as “Web-based services

---

\(^1\)www.sky.fm
\(^2\)www.ozsale.com.au
\(^3\)www.brandsexclusive.com.au
that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system (boyd & Ellison, 2007, p. 211).

In order to participate within an SNS, users are typically required to fill out their profile pages with basic and personal information such as full name, e-mail address, location, gender, date of birth, and profile picture. Users can then share their thoughts and activities with people they know online and in real life. Information shared on SNSs can be classified into five groups (Krishnamurthy & Wills, 2008) known as: (1) Thumbnail (A brief profile contains at least full name and image), (2) Greater profile (Additional information including interests and relationship status), (3) List of friends, (4) User generated content (e.g., pictures, video, links and comments), and (5) Comments (Status updates, testimonial and tags).

### 3.2.2 The use of users’ intentionally shared information

Information supplied by users can be used by the site owners for different purposes. Some information such as name, credit card details, postal address and e-mail address are necessary in processing payment and delivery of purchased items. Other details about users including location, gender or clickstream are necessary for the site to give users personalised contents or tailored advertisements (Krishnamurthy & Wills, 2006a; Fisher, 2009). For example, eBay searched items can be sorted by the closest distance or location based on the location details provided by users during registration.

Users’ intentionally shared information can be available or transferred to parties other than the sites visited by users. For example, information such as contact and credit card details supplied by customers who book flights or hotel rooms through Lastminute.com.au can be disclosed by Lastminute.com.au to third party suppliers such as flight companies or hotels. In addition, by entering into an online competition hosted by Lastminute.com.au (e.g., winning a return flight to Europe, or 5 night stays in Perth, Australia), users are required to provide information such as e-mail address, phone numbers, and postal address to Lastminute.com.au, and this information will be available to Lastminute.com.au’s third party suppliers or business partners that are unknown to users. The storage and process of these details are subject to those third parties’ terms and conditions and policies that are not controlled by Lastminute.com.au.

Information given and generated by SNS users can be used by SNSs for the purpose of connecting users with other people they may know (Facebook, 2010). For example, Facebook and Google Plus usually suggest people a user may know and may want to add as friends based
on their information including school, university, hometown or common friends. Meanwhile, information posted in Web forums or blogs can be accessed and used as resources or references by other Internet users. For example, Charles Sturt University provides internal forums that can be used by students and lecturers to exchange information and study resources.

Users’ intentionally shared information within an SNS is being shared by SNS to third party sites via the use of Application Programming Interface (API) (Ko, Cheek, & Shehab, 2010). The API services allow the third party sites to develop social applications without having to build their own social networks. These applications provide interesting contents to users’ existing profiles. For example, Trove, a Facebook third party application, allows Facebook users to customise and personalise news and other interesting stories aggregated online they like to see while using Facebook. In order to use this application, users are required to grant permission to Trove to access their information supplied to Facebook. This information includes public profile information (name, profile picture, gender, language, country and other public information), friend list, e-mail address, birthday, hometown, interests, current city, personal descriptions and Likes on Facebook, as shown in Figure 3.1. It is not clear how users’ information will be used by those parties.

![Trove Facebook third party application](image)

**Figure 3.1:** Trove Facebook third party application

Users’ intentionally shared information can also be available to audiences beyond users’ expectations. For example, Stephanie Rice, an Australian Olympic swimmer, was dismissed by her sponsor Jaguar over her Twitter message and was forced to apologise to those offended by her Tweet when she claimed the South African rugby team were homosexual (Bymes, 2010). Her gay friend, Matthew Mitchum, defended her, saying she meant no harm; but nevertheless, her message was offensive and thoughtless. A number of conclusions can be drawn from this situation: (a) Rice had no idea of who (in the world) has access to her Tweets, (b) Rice treated Twitter as a private space, and/or (c) such an incident would never happen if she said exactly the same thing with her close friends, in a private space (e.g., her home). In this case, SNS privacy settings are seen to be one of the many factors that make users’ information available to the public beyond their expected audience (Gross & Acquisti, 2005; Krishnamurthy & Wills, 2008).
Online users may open themselves to risks such as embarrassment, stalking, and identity theft which could harm them physically and mentally (Gross & Acquisti, 2005; boyd & Heer, 2006). SNSs are said to blur the line between professional and personal life (Barnes, 2006) when, for example, users’ colleagues or bosses befriend them on Facebook, where users post pictures or comments that should not be seen by their professional peers. Without proper privacy settings, those contents could also be seen by the public or other Facebook users.

In addition, the richness of personal information that includes physical location (e.g., home or university address), contact details (e.g., telephone number and e-mail address), and profile pictures could help a potential stalker to determine a user’s real-time location. Other additional information such as date of birth, hometown, mobile numbers, family tree details (e.g., cousin, brother or sister), and maiden name, can be exploited in the theft of identity. Identity theft is a fraud that involves gaining benefits by pretending to be someone else (ACCC, 2013). In days gone by, thieves searched rubbish bins or stole letters in order to gain personal information. Now, this type of information is readily available on SNSs.

3.3 Unintentionally shared information

3.3.1 An overview

The term “unintentionally shared information” refers to information about a user that can be gathered or leaked without a user’s intention to reveal or without a user’s awareness that information is being shared. This type of information may include users’ personally identifiable information (PII) and other information including their browsing activities or clickstream, search keywords, and technical information (e.g., IP address, and Internet access device unique identifier). PII refers to “information which can be used to distinguish or trace an individual’s identity either alone or when combined with other public information that is linkable to a specific individual” (Johnson-III, 2007).

The gathering and processing of this type of information has been increasingly outsourced to third parties while users may not be aware of the collection. As stated in Section 1.7, the term “first party sites” refers to sites that are directly or intentionally accessed by users, whereas third party sites are those which are not directly requested by users. Third party sites can be data aggregators or advertisers which display targeted advertising contents on the first party Web pages. For instance, Jane is searching for allergy tablets and nasal spray on a pharmacy Web site. She might not be aware that her searches can be collected by not only the pharmacy site (which is the first party site) but also by various third party sites such as advertisers and data
aggregators. Jane did not intend to disclose this type of information to third party sites. The collection of users’ clickstream data among Web pages is seen to be a mechanism to extract user-related information. This type of information helps to tailor advertisements and generate revenue for the site owners (Krishnamurthy & Wills, 2006a).

3.3.2 The collection of users’ unintentionally shared information

Many studies have demonstrated that pieces of users’ information including their online activities are being monitored and collected by numerous first and third party sites (Krishnamurthy & Wills, 2006a, 2006b, 2009; Soltani, Canty, Mayo, Thomas, & Hoofnagle, 2009; Krishnamurthy & Wills, 2010a, 2010b; Angwin, 2010; Ayenson, Wambach, Soltani, Good, & Hoofnagle, 2011; Wills & Zeljkovic, 2011; Krishnamurthy & Naryshkin, 2011). This type of information can be collected by a number of techniques including the traditional HTTP cookies, flash cookies and Web bugs or Web beacons. This is happening due to the extraneous objects or contents that are displayed on first party sites, and that are populated by third party sites or servers. Sometimes these contents are visible under the form of advertisements, while sometimes they are in the form of an image in 1x1 pixel size that is not visible by users.

HTTP Cookies

As mentioned in Section 1.7, HTTP or HyperText Transfer Protocol, is used to communicate between the browser, any intermediate machines, and Web servers (Comer, 2000). It is a stateless protocol, which means each request or response is treated independently. So in order to remember the state, it uses a small text file called an HTTP cookie, which is stored by the browser at the user’s machine (Kristol, 2001). A cookie is a piece of data given by the server to the client, stored by the client, and returned to the server when the client returns. It was first developed in 1994 for the purpose of assisting users in online shopping by serving as a virtual shopping trolley (Hormozi, 2005).

Cookies were not intended to be used as spying mechanisms, but rather for the purpose of enhancing a user’s interaction with the Web by informing the Web server that the same user had returned (Sipior, Ward, & Mendoza, 2011). The Web server recognises the user by accessing the information previously stored in the cookie file, and the user can resume interaction where it had been left off on their previous visit. The HTTP cookie is not a program, but it can be used by marketers or Web developers to store information about online users’ interests or searches (Hormozi, 2005). For example, when a user browses for kitchen equipment on an online store Web page, that store can save information about the items browsed and/or added
to the shopping trolley by the user. The store can suggest other equipment or items that might be of interest to that user.

**Flash cookies**

Users’ information can also be collected by flash cookies which are stored and accessed by Adobe Flash (Herman, 2010). While HTTP cookies can be erased anytime and cannot be relied on for saving important information about users, flash cookies, on the other hand, are more persistent (Soltani et al., 2009). Flash cookies protect other cookies from being deleted by making backup copies, and they are loaded on more than 98% of computers (Marshall, 2005). They are stored in a different location to HTTP cookies, can contain up to 100KB of information and do not have any expiry date by default (Brinkman, 2007; Soltani et al., 2009).

Similarly to HTTP cookies, flash cookies were originally used to enhance users’ interaction and browsing experience with a Web server by remembering the state (Soltani et al., 2009; Sipior et al., 2011). They can be used for basic functions such as saving the level of a user’s volume, language preferences, or where a user left off playing a flash video. Flash cookies create uncertainty for online browsing because they are not controlled by the browser. Hence, clearing histories, cache and cookies in the browser does not remove flash cookies. In addition, flash cookies can be used to track users’ movements across different browsers and browsing sessions (Soltani et al., 2009).

**Web beacons**

Web beacons, beacons, Web bugs, or clear gifs present themselves in graphic formats of 1x1 pixel in size, tagged in HTML documents placed on a Web site or an e-mail message to monitor user behaviour (Smith, 1999). While users can choose to accept or deny cookies, they may not be able to detect beacons because they are placed within an HTML file as small, transparent, and invisible files. Beacons can monitor and record a user’s typed entries and mouse movements on a page (Angwin, 2010). They can be used by advertising networks to harvest information about online users’ clickstreams or browsing behaviours, and add that information to personal profiles which are identified by the browser cookies (Smith, 1999).
3.3.3 The use of users’ unintentionally shared information

Similar to users’ intentionally shared information, their unintentionally shared information can be used by both first and third party sites for different purposes. The collected unintentionally shared information can be used by first party sites to analyse their network traffic in order to operate, maintain, and improve their Web sites, products and services. This information can also be used to offer personalised contents and search results based on users’ browsing interests, preferences and experiences (9MSN, 2012; Speedo, 2012). First party sites’ owners increasingly use third party analytics tools in order to evaluate traffic or performance on their own sites (Tene & Polonetsky, 2012). Therefore, users’ information can become available to those third party sites.

The number of third party sites used by first party sites has been increasing significantly from 40% in October 2005 to 70% in September 2008 (Krishnamurthy & Wills, 2006b, 2009). In addition, user information available to third party sites became more personal and identifiable to a specific user (e.g., a user’s full name, e-mail address, and location) when SNS emerged (Krishnamurthy & Wills, 2010a, 2010b). Due to the low and decreasing costs of storing information in digital format, it is possible to continuously monitor and record users’ online movements and information over time. It is not clear how this type of information is or will be used or stored by third party sites, apart from for their advertising purposes.

Information about users including their personal information, past and current searches, interests, or clickstream, can be used by marketers or advertisers for the purpose of online advertising (Tene & Polonetsky, 2012; Wills & Tatar, 2012). Online advertising revenues have kept growing and set new high at $9.6 billion in the first quarter of 2013 (IAB, 2013). According to Mane, “Internet advertising revenue continues to exhibit double-digit growth, even as the business matures... This is an accomplishment that can be attributed to growing recognition by marketers that digital advertising is a critical part of all marketing in today’s world” (IAB, 2013).

Users’ unintentionally shared information is also being used for the purpose of targeted advertising, a technique which has become very popular (Yang, Dia, Cheng, & Lin, 2006) due to the fact that advertisers are able to provide the right advertisements for the right consumers based on different traits such as consumers’ searches, viewed content or behaviours online. Meanwhile, content-dependent advertising or contextual advertising is a type of targeted advertisement in which advertising contents are selected and served to consumers based on the Web content (such as keywords on the page) viewed by users, and displayed to users on a banner or pop-up advertisements (Yang et al., 2006; Kim, 2010). For example, when a user is viewing a website pertaining to cooking, the user may see relevant advertisements such as sales
or discounts on cooking utensils or cookbooks. Unintentionally shared information can also be gathered and used for the purpose of behavioural advertising or behavioural targeting. This is another form of targeted advertising in which advertisement contents displayed to users are based upon their behaviour online or their specific behavioural history (Yan et al., 2009; Barocas & Nissenbaum, 2009; McDonald & Cranor, 2010b). Users’ online behaviours are represented by their Web browsing behaviour and their search queries. Unlike contextual advertising technique that scans the text of a website for keywords, behavioural advertising uses various data mining techniques that process information about users in order to yield predictive models that helps in advertisement decisioning and targeting systems. The best example to illustrate this would be the case of the researcher’s searches. The researcher personally spent a period of time doing online searches to gather the facts about weight loss. The advertisements about diets, diet plans, exercises, or slim shakes, appeared across different Web sites and on different computers, visited and used by her for a long period of time.

3.4 Unintentionally shared information and related works

Earlier works focused on the factors or extraneous contents that may affect the performance, and eventually the popularity of a Web site or service (Krishnamurthy & Wills, 2000; Krishnamurthy, Wills, & Zhang, 2001; Liston & Zegura, 2001; Bent & Voelker, 2002). Those extraneous contents included advertisement banners, pop-up windows, animations, cookies and scripts. It was later found that those extraneous contents were present on the majority of popular Web sites (Krishnamurthy & Wills, 2006a), and that this presence kept growing significantly (Krishnamurthy & Wills, 2009). Through the examination of HTTP headers (see definition in Section 1.7), those extraneous contents were also found to be used among Web sites as mechanisms to share and collect users’ unintentionally shared information including personal details such as name and e-mail address, and browsing behaviours (Krishnamurthy & Wills, 2009, 2010a; Angwin, 2010; Mayer, 2011; Krishnamurthy & Naryshkin, 2011).

Bits and pieces of users’ information are being aggregated as a result of browsing seemingly unrelated Web sites or downloading extraneous contents from third party site servers (Krishnamurthy & Wills, 2006b). Those third party sites are found to be able to track users’ online movements across different sites by using HTTP cookies. For example, when a user visits a number of Web sites that retrieve objects like advertisements from the same third party server, that third party server is able to associate that user with the requests from his/her visited pages via the use of cookies. The top third party sites are AOL, Google, DoubleClick (a Google
franchise), Microsoft, Omniture, ValueClick and Yahoo (Krishnamurthy & Wills, 2009).

Let us consider the HTTP conversation between the researcher’s browser and Google illustrated in Table 3.1.

- The researcher types www.google.com.au into her browser. The browser sends a HTTP request to the Google server for the content to display on the browser as shown in Table 3.1(a). The HTTP header field, User-Agent, specifies that the researcher uses Firefox browser on her Mac OS computer to access the Internet.
- Google then replies back to the browser with a HTTP response message as shown in Table 3.1(b). Since HTTP is a stateless protocol, Google sets cookies to the researcher’s computer in order to remember the state with Set-Cookie. There are two cookies: PREF that will expire on 09-Aug-2015, and NID that will expire on 08-Feb-2014.

<table>
<thead>
<tr>
<th>(a) Client or HTTP request</th>
</tr>
</thead>
<tbody>
<tr>
<td>GET / HTTP/1.1\r\n</td>
</tr>
<tr>
<td>Host: <a href="http://www.google.com.au%5Cr%5Cn">www.google.com.au\r\n</a></td>
</tr>
<tr>
<td>User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10.8; rv:22.0) Gecko/20100101 Firefox/22.0\r\n</td>
</tr>
<tr>
<td>Accept: text/html,application/xhtml+xml,application/xml;q=0.9,<em>/</em>;q=0.8\r\n</td>
</tr>
<tr>
<td>Accept-Language: en-US,en;q=0.5\r\n</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(b) Server or HTTP response</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTTP/1.1 200 OK\r\n</td>
</tr>
<tr>
<td>Date: Fri, 09 Aug 2013 00:43:46 GMT\r\n</td>
</tr>
<tr>
<td>Expires: -1\r\n</td>
</tr>
<tr>
<td>Cache-Control: private, max-age=0\r\n</td>
</tr>
<tr>
<td>Content-Type: text/html; charset=UTF-8\r\n</td>
</tr>
<tr>
<td>Set-Cookie: PREF=VALUE1; expires=Sun, 09-Aug-2015 00:43:46 GMT; path=/; domain=.google.com.au\r\n</td>
</tr>
<tr>
<td>Set-Cookie: NID=VALUE2; expires=Sat, 08-Feb-2014 00:43:46 GMT; path=/; domain=.google.com.au; HttpOnly\r\n</td>
</tr>
</tbody>
</table>

Table 3.1: HTTP Conversation between the browser and Google

As briefly noted in Section 1.2 of Chapter 1, previous studies that examined the HTTP header have shed light on the current practice of sharing and gathering user information. Those studies investigated a large number of sites, including SNSs and other sites across different categories (Krishnamurthy & Wills, 2008, 2009, 2010a; Mayer, 2011). With the emergence and popularity of SNSs, third party sites, which are advertisers, are seen to be able to not only infer information about users, but also directly obtain identifiable information about users (Krishnamurthy & Wills, 2008, 2009, 2010a, 2010b; Mayer, 2011). Different categories of sites (e.g., online store, online news) are seen to leak or transfer users’ browsing details or clickstream, which is not identifiable to a user, to third party sites which are advertisers. On the other hand, SNSs are seen to leak or transfer users’ identifiable information to third party sites. This information includes a users’ name, SNS unique identifier, location, e-mail address, and activities within SNSs (Krishnamurthy & Wills, 2010a). Users’ mobile device unique identifier and location are also available to third party sites from mobile SNSs (Krishnamurthy & Wills, 2010b).
These studies did not examine the information sharing in the HTTP headers from real life online users who browse different categories of sites (Purcell, 2011; Zickuhr & Smith, 2012). Some users are also privacy savvy and use protection on their browsers for private browsing mode (Bursztein, 2012). In addition, it is important to examine the information sharing among different categories of sites because, while the information leaked from SNSs tends to be more personal and identifiable to a specific user, the combination of this type of information and the browsing behaviours among other sites could reveal much about a person’s life. Hence, there is a need to include real life online users to better reflect online browsing activities.

### 3.5 Need for this stage of research

This research focused only on information about online users that is not intentionally shared or disclosed by them, yet is available to third party sites that have not been directly visited by users. The sharing of users’ unintentionally shared information from first party to third party sites can be sophisticated and complicated and could be happening without users’ awareness. Privacy-conscious users may choose to share less or may be cautious with what they disclose on their SNS profiles, blogs, or Web forums, to the audience or other Internet users. However, they may not be aware that their online activities and other pieces of information can be gathered by not only first party, but also third party sites.

According to the rationale mentioned earlier in Section 3.4, there is a need to investigate the sharing of users’ unintentionally shared information in the HTTP headers from the angle of real life online users. It is important to involve online users in the investigation because the browsing activities and sites can be diverse among different individuals. The investigation of this diversity may yield different outcomes that were not reported in the existing literature. There is also a need to involve online users from Cambodia where ICT is very under-researched and where users’ browsing behaviours and awareness of online tracking could be very different from those in the developed countries like Australia and the United States. The researcher hopes that this study addresses a gap in the literature and improves understanding of this under-researched society.

In addition, as alluded to in Chapter 1, the second stage of this thesis also examined participants’ perception and awareness of information sharing in HTTP headers (that resulted in online tracking) and privacy online. Hence, the outcomes from Chapter 4 were also necessary for the researcher to develop and design the interview questions with participants. It also helped to shape the focus of the whole thesis - the collection of users’ unintentionally shared information.
3.6 Research intent and design

Based on the rationales given in Section 3.4 and Section 3.5, this section of the research intended to investigate information sharing and gathering in the HTTP headers among online users from Cambodia. It aimed to address the following questions:

1. What is the nature of information sharing in the HTTP headers?
   (a) What type of information is being shared?
   (b) With whom is this information being shared?
   (c) What are the implications of this sharing?

In order to address these questions, this research first employed an exploratory study in which the researcher investigated the HTTP messages resulting from browsing activities. The exploratory study helped the researcher to establish what types of devices, applications, and networks were needed to carry out larger scale experiments. It also allowed the researcher to assess and mitigate the risks associated with recording and analysing the research participants’ browsing activities; for example, there was a risk that the password could be revealed if it was not encrypted.

The exploratory study was conducted by having the researcher browse online while having these activities recorded and analysed. The browsing and recording followed a specific set of protocols and procedures, as well as chosen sites and activities. The details of the process and analysis of the outcome are presented in Chapter 4. Once the feasibilities and requirements had been assessed in the exploratory study and ethics approval had been obtained from Charles Sturt University (Appendix A), larger scale experiments were conducted with participants in Cambodia. Chapter 5 presents all the steps and processes as well as the outcome of the analysis.

This study is similar in some ways to those by Krishnamurthy and Wills (2010a); Mayer (2011) and Krishnamurthy and Naryshkin (2011) which investigated information leakages in HTTP headers. The exploratory study is also similar to studies by Dwyer (2009), and Wongyai and Charoeunwatana (2012) as it takes the form of a small case study. However, this study differs from others in terms of categories of sites and online activities. This study did not investigate large numbers of sites or large number of SNSs alone; nor did it focus on only one organisation (Dwyer, 2009) or site (Wongyai & Charoeunwatana, 2012). Rather, the exploratory study examines a small number of sites frequently visited by most users while employing different sets of browsing activities and Google search trends that are common among most people (details in Chapter 4), whereas the sites and activities in the experiments conducted in Cambodia were chosen by the participants themselves.
Chapter 4

Users’ unintentionally shared information in the HTTP headers: an exploratory study

This chapter provides an account of the exploratory study conducted by the researcher in order to determine the appropriate tools and method to collect the data from research participants in Cambodia. It also presents the findings of the experiments. The chapter is organised as follows. Section 4.1 details the justification and choice in selecting sites and online activities to be performed within the study. The steps and processes in recording, collecting and analysing the data are presented in Section 4.2, while the findings of this study are summarised in Section 4.3. Section 4.4 discusses the findings and implications and Section 4.5 concludes the exploratory study.

4.1 Online activities, sites, and searches used in the study

This exploratory study was intended to provide the researcher with a better understanding of the nature of information sharing in the HTTP headers among Web sites. It also allowed the researcher to test the feasibility of undertaking a more extensive study, and to develop the methods to be employed in larger scale experiments. At the same time, it also helped the researcher to address the following research questions:

- What is the nature of information sharing in the HTTP headers?
  1. What type of information is being shared?
  2. With whom is this information being shared?
3. What are the implications of this sharing?

In order to closely reflect online users’ online activities, the researcher decided on the set of online activities to be performed based on the report by Pew Internet (Purcell, 2011) and selected the sites associated with chosen activities based on the ranking in Alexa\(^1\) in conjunction with personal browsing behaviours. The top or most popular online activities surveyed by Pew Internet (Purcell, 2011) were: checking e-mail messages, using SNSs, doing online shopping, reading online news articles, and performing online searches (e.g., by using Google search engine). The researcher also chose popular search trends based on Google statistics (Google, 2012c). After choosing top visited sites (ranked by Alexa) combined with popular online activities (Purcell, 2011), and Google search trends (Google, 2012c), in conjunction with the researcher’s online browsing preferences, a list of sites can be summarised in Table 4.1.

<table>
<thead>
<tr>
<th>E-mail</th>
<th>SNS</th>
<th>Online shop</th>
<th>News</th>
<th>Google search</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yahoo</td>
<td>Facebook</td>
<td>eBay</td>
<td>9 MSN news</td>
<td>Lynco007</td>
</tr>
<tr>
<td>Gmail</td>
<td>Twitter</td>
<td></td>
<td>ABC News</td>
<td>Taste.com.au</td>
</tr>
<tr>
<td></td>
<td>LinkedIn</td>
<td></td>
<td></td>
<td>Weatherzone</td>
</tr>
<tr>
<td></td>
<td>YouTube</td>
<td></td>
<td></td>
<td>Wikipedia</td>
</tr>
</tbody>
</table>

Table 4.1: Online activities and sites chosen for the studies

Wireshark\(^2\) was chosen as a tool to record and collect the data. It is a network protocol analyser that captures and displays all HTTP traffic, such as communication between the browser or application and the requested sites or servers (Figure 4.1). The rationale for choosing Wireshark is that the researcher could triangulate the data by conducting the experiments across different operating systems (OS) and different network environments (with and without proxy settings). Unlike Fiddler (http://fiddler2.com/), Wireshark is more versatile and can be installed on any OS, which will be suitable for this study given the participants will have different devices with different OSes. Wireshark can be used to record the activities of any application or browser running on the device. Each recorded performance can be saved or exported into different file formats for later analysis. The only drawback is that encrypted information or information transmitted over SSL or HTTPS packets are not observable. However, as the aim was to analyse HTTP packets, this encrypted information has no effect on this study.

---

4.2 Data collection and analysis

The study included multiple browsing sessions across different computers and OSes. Each session involved the researcher performing a set of online activities (as summarised in Table 4.1), while the activities were recorded by Wireshark. The experiments were conducted on Windows, Linux and Mac OS machines using the Firefox browser, with and without the AdBlock Plus Firefox extension\(^3\), over two separate networks (with and without a proxy server). Users are often encouraged to create accounts for many categories of sites they visit. However, for this study, the researcher already owned accounts for the sites under investigation, so this study did not investigate the possible information leakage or sharing during the sign-up process.

Examples of the researcher’s online activities included, but were not limited, to the following:

- **E-mail**: signing in, checking, reading and sending e-mail messages.
- **SNSs**: signing in, checking her and her friends’ profiles, checking messages, where feasible, playing Farmville third party application and clicking on the advertisements.
- **Online shopping**: signing in, searching for and occasionally purchasing a few items (e.g.,

---
\(^3\)AdBlock Plus can be installed as an extension or add-on in various browsers such as Firefox, Google Chrome and Opera. It prevents Web pages from downloading and displaying extra contents like advertisements.
• News: browsing different types of articles (e.g., technology, health, or national news).
• Online search: checking the daily weather forecast for current location, searching for some general knowledge about a specific topic (e.g., hay fever).

Each experiment or browsing session included the following steps:

• Making notes about which browsing activities will be performed.
• Terminating other applications running on the device which may also be using the HTTP protocol.
• Opening the browser and clearing all the cookies and search histories.
• Running Wireshark and starting to record the HTTP messages.
• Performing a set of actions as planned (e.g., those noted in the list above).
• Taking notes of which exact actions are performed if there are changes.
• Stopping the recording and saving the trace when the browsing actions are completed.
• Examining and observing the saved HTTP messages line by line and taking notes of what/who the third party sites are and what types of information are being shared to them.

The researcher examined the HTTP headers in each file (as shown in Figure 4.1) resulting from the online browsing in order to identify the types of information being shared and the types of third party sites involved in this sharing. Let us examine an HTTP conversation between the browser, first party and third party sites in Table 4.2.

(a) The researcher visits the Mediterranean chicken pasta salad recipe page on taste.com.au. The browser makes an HTTP request to taste.com.au in order to retrieve the page content.

(b) taste.com.au receives the HTTP request and returns an HTTP response which contains the page content (in HTML format) as well as Javascript code. The browser executes the code that it requires to fetch an image content from another site, imrworldwide.com.

(c) The browser then sends another HTTP request to imrworldwide.com for the image content to display on the recipe page. Within this request, taste.com.au is seen to leak the researcher’s search keyword, Mediterranean chicken pasta salad to a third party site, imrworldwide.com, in the HTTP header.

(d) imrworldwide.com then sends an HTTP response (contains 1x1 pixel image) and sets cookies in the browser.
4.3 Findings

Visits to a few Web sites result in thousands of lines of HTTP messages captured by Wireshark, whether on Windows, Linux or Mac OS. Similar to the findings reported by Krishnamurthy et al. (2007), browser extensions help to limit the number of third party site connections, but does not eliminate them. Not all the advertisements were removed from the visited pages while AdBlock Plus was enabled. The researcher summarised the findings by first reporting on the level of information propagation going beyond one level of the third party sites which was not covered in the previous literature, then on the types of third party site being contacted at each propagation level, while also pinpointing the types of information being transferred from the first party site. It was found that the researcher’s identifiable and non-identifiable information was being shared to third party sites while she was visiting first party sites. Third party sites were identified to be advertisers, data aggregators, and SNSs (e.g., Facebook, Twitter, and Google Plus).
4.3.1 First level traverse: From first party sites to third party sites

Third party sites which are advertisers or data aggregators

It is not surprising and appears to be consistent with the literature that first party sites usually transferred user’s search keywords to third party sites which are advertisers or data aggregators. The connection is necessary to fetch advertisement contents from advertiser servers to display on first party site pages. In this case, for example, paying a single visit to some sites such as taste.com.au (for a vanilla cupcake recipe), Lyrics007 (for a song lyric: *Yesterday* by the Beatles), and Nine News (for an article: *Google gets personal with search results*), resulted in at least 10 connections to third party sites while also transmitting the researcher’s searches to those sites (Table 4.3).

<table>
<thead>
<tr>
<th>Taste.com.au (Recipe: vanilla cupcakes)</th>
<th>Lyrics007 (Song title: Yesterday)</th>
<th>Nine News (Article: Google gets personal with search results)</th>
</tr>
</thead>
<tbody>
<tr>
<td>news-statistics.com</td>
<td>clickfuse.com</td>
<td>shared.9msn.com.au</td>
</tr>
<tr>
<td>sops.news.com.au</td>
<td>addthis.com</td>
<td>msnportal.112.2o7.net</td>
</tr>
<tr>
<td>trakr-news.com.au</td>
<td>fastclick.net</td>
<td>inmrworldwide.com</td>
</tr>
<tr>
<td>google-analytics.com</td>
<td>google syndication.com</td>
<td>n.adxns.com</td>
</tr>
<tr>
<td>notebookmagonzin.com</td>
<td>ringtonemaker.com</td>
<td>bs.serving-sys.com</td>
</tr>
<tr>
<td>fashion.vogue.com.au</td>
<td>b.scorecardresearch.com</td>
<td>b.scorecardresearch.com</td>
</tr>
<tr>
<td>doubleclick.net</td>
<td>apmbf.com</td>
<td>widget.twimg.com</td>
</tr>
<tr>
<td>bs.serving-sys.com</td>
<td>rtbiddler.net</td>
<td>facebook.com</td>
</tr>
<tr>
<td>jdn.monster.com</td>
<td>amazonaws.com</td>
<td>twitter.com</td>
</tr>
<tr>
<td>facebook.com</td>
<td>advertising.com</td>
<td></td>
</tr>
<tr>
<td>twitter.com</td>
<td>abmr.net</td>
<td></td>
</tr>
<tr>
<td>getprice.com.au</td>
<td>googleapis.com</td>
<td></td>
</tr>
<tr>
<td>imrworldwide.com</td>
<td>doubleclick.net</td>
<td></td>
</tr>
<tr>
<td>unica.com</td>
<td>facebook.com</td>
<td></td>
</tr>
<tr>
<td></td>
<td>api.google.com</td>
<td></td>
</tr>
<tr>
<td></td>
<td>platform.twitter.com</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3: List of third party sites connected to by the browser while visiting first party sites

As shown in Table 4.4, there were two cases of information sharing for Facebook. In the first case, connections were made only to Facebook CDN (Content Delivery Network) if no third party applications or advertisements on Facebook were clicked or used. In the second case, Facebook shared the researcher’s unique ID to Zynga (the online game company which hosts the Farmville application) when the researcher played Farmville from Facebook. Multiple requests started to be forwarded by Farmville to third party sites such as DoubleClick. However, transmission of researcher’s Facebook ID from Farmville to advertisers was not observed. Meanwhile, although there were connections to third party sites, leaks or shares of identifiable information from Twitter were not observed, although Twitter page movements were detected. LinkedIn, on the other hand, was seen to share the researcher’s information (in this case: LinkedIn ID and full name) to third party servers, like DoubleClick (Table 4.5).
Table 4.4: List of third party sites connected to by the SNS

<table>
<thead>
<tr>
<th>LinkedIn (No click on ads)</th>
<th>Twitter (No click on ads)</th>
<th>Facebook (Farmville application)</th>
<th>Facebook (No click on ads)</th>
</tr>
</thead>
<tbody>
<tr>
<td>google-analytics.com</td>
<td>google-analytics.com</td>
<td>akamaid.net</td>
<td>No third party sites</td>
</tr>
<tr>
<td>imrworldwide.com</td>
<td>twing.com</td>
<td>googletagservices.com</td>
<td></td>
</tr>
<tr>
<td>quantserve.com</td>
<td></td>
<td>quantserve.com</td>
<td></td>
</tr>
<tr>
<td>b.scorecardresearch.com</td>
<td></td>
<td>doubleclick.net</td>
<td></td>
</tr>
<tr>
<td>doubleclick.net</td>
<td></td>
<td>rubicomproject.com</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>googlegadservices.com</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>rtbidder.net</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>socialvi.be</td>
<td></td>
</tr>
</tbody>
</table>

GET /h?...
Host: ad.au.doubleclick.net
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10.6; rv:8.0.1) Gecko/20100101 Firefox/8.0.1
Referer: http://www.linkedin.com/profile/... a user’s LinkedIn ID and full name

Table 4.5: The leakage of the researcher’s LinkedIn ID and full name

Third party sites which are SNSs

The researcher’s information and searches were transferred from first party sites to third party SNSs. New SNS widgets, such as Facebook’s Like button, Twitter’s Tweet button and Google’s +1 button, enabled site users to share content from other Web sites with their SNS friends (Facebook, 2012c; Twitter, 2012a; Google, 2012a). Facebook’s Like button also enabled site owners to have a view of the number of likes on their domain both daily and demographically. It was observed that when the researcher visited sites embedded with those widgets, with or without logging into any SNS sites, those first party sites always sent HTTP requests to the SNS to populate the page with the SNS buttons. As shown in Table 4.3, Taste.com.au, Lyrics007, and Nine News, connected to at least three SNSs - Facebook, Twitter and Google Plus - because SNS widgets reside on those first party sites. For example, Table 4.6 shows Lyrics007 sending a request to Twitter to retrieve the widget while also sharing information about the researcher’s currently visited page to Twitter. Twitter recognises the same visit from the cookies that were set when the researcher logged into Twitter.
GET /widgets.js HTTP/1.1
Host: platform.twitter.com
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10.6; rv:8.0.1) Gecko/20100101 Firefox/8.0.1
...
Referer: http://www.lyrics007.com/the song title and the artist name
Cookie: k=Twitter cookie 1

Table 4.6: Lyrics007 shares the researcher’s search keyword to Twitter while retrieving Twitter widget

4.3.2 Second level traverse: From third party sites to other third party sites

It was observed that the researcher’s information and browser connection propagated beyond one level from the first party sites. The following example illustrates this finding.

eBay:

When the researcher visited eBay and searched or purchased an item, eBay shared the search keyword with other third party sites, such as DoubleClick. DoubleClick was also seen to share that information with other third party sites, like amgdgt.com and b.scorecardresearch.com (Table 4.7). Information was seen to be transmitted from eBay to a third party site, and from that third party site to another third party site as shown in Figure 4.2.

Table 4.7: The traverse of eBay the researcher’s search keyword

Lyrics007:

Lyrics007 was similar to eBay in that there was a second level of information sharing to other third party sites. For example, the song title propagated from Lyrics007 to jangonetwork,
then from jangonetwork to DoubleClick (Table 4.8). In addition, when the researcher visited Lyrics007 (1), the browser connection went to rubiconproject.com (2), then from rubiconproject.com to w55c.net (3), and from w55c.net to bluekai.com (4) (Figure 4.3). However, information sharing ended in the second level as shown in Figure 4.3.

Table 4.8: Lyrics007 traverse

<table>
<thead>
<tr>
<th>Step</th>
<th>Request Details</th>
</tr>
</thead>
</table>
| 1    | GET /007... "song title and artist name"  
Host: jmn.jangonetwork.com  
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10.6; rv:8.0.1) Gecko/20100101 Firefox/8.0.1  
Referer: http://www.lyrics007.com/"the song title and the artist name" |
| 2    | GET /widgets.js HTTP/1.1  
Host: partner.googleadservices.com  
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10.6; rv:8.0.1) Gecko/20100101 Firefox/8.0.1  
Referer: http://jmn.jangonetwork.com/ "song title, and artist name" |
| 3    | GET /gampad/ads?...  
Host: pubads.g.doubleclick.net  
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10.6; rv:8.0.1) Gecko/20100101 Firefox/8.0.1  
Referer: http://jmn.jangonetwork.com/"song title, and artist name" |

4.4 Discussion

In this exploratory study, it was observed that third party sites collected the researcher’s unintentionally shared information, both identifiable and non-identifiable. Those third party sites were identified to be advertisers or data aggregators and SNSs such as Facebook, Twitter, and Google Plus (Table 4.9). The researcher’ information propagated from first party sites (both
Figure 4.3: The traverse of the browser’s connection

SNS and non-SNS) to third party sites, and also from third party sites to other third party sites. If first party sites use or connect to the same third party sites, those third party sites are able to track online movements across different sites, among SNS and non-SNS, via the use of HTTP cookies. The tracking can be classified into two categories: the tracking by third party sites which are advertisers or data aggregators, and the tracking by third party sites which are SNSs.

<table>
<thead>
<tr>
<th>First party sites</th>
<th>The leaked information</th>
<th>Third party sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yahoo</td>
<td>User’s clickstream within Yahoo</td>
<td>Advertisers or data aggregators</td>
</tr>
<tr>
<td>Gmail</td>
<td>Not observable</td>
<td>Not observable</td>
</tr>
<tr>
<td>Facebook (no ads or app clicked)</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Facebook (e.g., Farmville app)</td>
<td>Facebook ID</td>
<td>Advertisers or data aggregators</td>
</tr>
<tr>
<td>Twitter</td>
<td>User’s searches and clickstream</td>
<td>Advertisers or data aggregators</td>
</tr>
<tr>
<td>LinkedIn</td>
<td>User’s clickstream and full name and ID</td>
<td>Advertisers or data aggregators</td>
</tr>
<tr>
<td>Taste.com.au</td>
<td>User’s searches and clickstream</td>
<td>SNSs and Advertisers or data aggregators</td>
</tr>
<tr>
<td>9 MSN News</td>
<td>User’s searches and clickstream</td>
<td>SNSs and Advertisers or data aggregators</td>
</tr>
<tr>
<td>eBay</td>
<td>User’s searches (e.g., hairclips)</td>
<td>Advertisers or data aggregators</td>
</tr>
<tr>
<td>Wikipedia</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>YouTube</td>
<td>User’s searches and clickstream</td>
<td>Google</td>
</tr>
<tr>
<td>Weatherzone</td>
<td>User’s location and clickstream</td>
<td>SNSs and Advertisers or data aggregators</td>
</tr>
</tbody>
</table>

Table 4.9: Summary of the information leakages in the study
4.4.1 Tracking by third party sites which are advertisers or data aggregators

Case: Scorecard Research

Scorecard Research⁴ is one of the most contacted third party sites among those visited sites in this study. They are able to obtain both the researcher’s identifiable information (in this case from LinkedIn: name and ID) and non-identifiable information (her searches). Therefore, within just one browsing session, Scorecard Research is able to associate the researcher with her online movements. In this case, Scorecard Research knows the researcher’s name from LinkedIn, knows that the researcher likes cooking (from her search for recipes) and listening to the Beatles.

4.4.2 Tracking by third party sites which are SNSs

Through the use of the SNS widgets, it has been shown that SNSs are able to track the researcher’s online movements, not only within SNSs but also across the many other sites that embed SNS widgets. SNSs themselves hold large amounts of personal information about a person’s life. There are two separate situations here. First, without logging into any SNS during a browsing session, SNSs can track the researcher’s movements without being able to link these activities to her profiles. However, in the second situation, if the researcher has remained logged-in to any SNS account (Facebook, Twitter, or Google Plus, or even Gmail or YouTube) while also browsing different Web sites, that SNS is able to combine her browsing activities with her personal profile, within that browsing session.

Case: Twitter

When the researcher logged into Twitter, Twitter set cookies to remember the state (e.g., a user’s credentials). The cookies stored information about the researcher, including her Twitter ID, name, and IP address. When the researcher visited other sites that embedded the Tweet button, Twitter associated the visits with the same cookie value set while logging in the account. Even after the researcher logged out of her Twitter account, Twitter was still able to track her movements outside Twitter, as the cookie values remain the same.

⁴https://www.scorecardresearch.com/Home.aspx
Case: Facebook

Facebook, on the other hand, is slightly different to Twitter in the sense that once the researcher logged out of Facebook, the cookies associated with her account (which include a user name, e-mail address and ID) were destroyed. However, other cookies (e.g., guest cookies) which are not directly associated with a particular person remain. Those cookies can still associate browsed Web sites with browsing devices (e.g., OS, browser, and IP address).

Case: Google

Google was observed to act differently in this study. The traffic associated with Gmail or Google Plus could not be observed in HTTP messages. Rather, Google uses another protocol called TLS (Transport Layer Security) which is a cryptographic protocol providing secured communication over the Internet. However, the tracking was observed when the researcher used the Google search engine or YouTube. The majority of the first party sites within this study usually directly connected to Google or its other domains (e.g., googlesyndication or googleanalytics) or its franchise (DoubleClick), so Google has the ability to track a user across different sites from the use of the same cookies.

Again, without logging into Gmail or Google Plus, there was no linkage of the researcher’s online search with a specific identifier. However, logging into either of these services caused the researcher’s online activities (among first party sites which use Google’s +1 buttons, sites discovered through Google search, or sites using Google ads) to be linkable with her Google profile via the HTTP cookie. The researcher also observed that once she signed into her Gmail account, she became automatically logged into other accounts such as Google Plus and YouTube. Though Gmail and Google Plus traffic data is not observable via the HTTP headers, the rest of the traffic via HTTP messages shows that Google can gain access to many details about a person’s life if that user logs into their Gmail account and stays logged in during the browsing session.

4.5 Conclusion

Within this study, while non-SNSs share or leak non-identifiable information, they reveal the researcher’s browsing habits or searches to third party sites (e.g., advertisers and/or SNSs). SNSs like LinkedIn and Facebook (when the researcher clicks on advertisements or third party applications) share the researcher’s identifiable information (LinkedIn case: her user name and ID, Facebook case: her Facebook ID) to third party sites. The traffic between the browser and
the Google Gmail and Google Plus services was not visible in the HTTP protocol because it is encrypted. However, based on Google HTTP cookies, it was observed that the researcher’s online movements can be tracked and linked to her Google profile by Google if she happens to log into any of its services (either Gmail, Google Plus, or YouTube) within the same browsing session.

It was also observed that the researcher’s information and the browser connection propagated to more than just one level, traversing from first party sites to third party sites, and from those third party sites to other third party sites. This type of propagation has not yet been discussed in the literature. Third party sites found in this study are advertisers, data aggregators and SNSs (such as Facebook, Twitter, and Google Plus). These third party sites are able to track the researcher’s movements across different sites by the use of cookies and SNS widgets. With a piece of identifiable information from a SNS (as in the case of LinkedIn: name and ID), the browsing habits or behaviours can be linked to a specific person.

Technologies that enable tracking, like HTTP cookies, are not new. Users’ online movements, as well as browsing behaviours, have been tracked for as long as cookies have existed. It becomes problematic when a specific user can be identified. If users never clears their browser histories and cookies, browsing profiles may include their online activities for 365 days of the year (some cookies can live up to 100 years). It appears that whenever users visit any Web site, it is inevitable that they will leave their digital footprints within and across sites.

These results provide insight into information leakages and the nature of behavioural tracking within a browsing session, where the researcher browsed both SNSs and non-SNSs frequented by many, and whose activities are common among many online users. Although the findings of this study are not intended to generalise about information gathering online, within just one browsing session, the combination of leaked information reveals much about the researcher. Information was propagated to more than just one level of third party sites, and those sites were able to track the researcher’s browsing habits and combine those details with her identifiable information.

Among these findings, the tracking by third party sites as SNSs appears to be problematic as Facebook, Twitter and Google Plus acquired information about the researcher’s online activities in addition to storing her personal information and activities on SNS profiles. The researcher would like to determine whether other online users, particularly users from Cambodia, are aware of this practice. In addition, she researcher would like those users to see the tracking ramifications resulting from their own online browsing prior to gauging their awareness and perception of this tracking.

This exploratory study demonstrated that experiments on a larger scale can be conducted with
20 research participants. Wireshark captures and reveals data that are not encrypted; hence, the researcher would be cautious about the possibility that participants’ passwords could be revealed if they are not encrypted. Wireshark worked well across different OSes and browsers, and proved to be an appropriate recording application for participants. Since the tracking by SNSs appears to be problematic, the researcher decided to focus this study on this type of tracking and associated impacts. Interview questions in the second stage of the research and the ethical analysis in the third stage of the research were built upon this focus. The next chapter, Chapter 5 provides details about Wireshark experiments with 20 participants in Cambodia.
Chapter 5

Information sharing in the HTTP headers: 20 participants

Chapter 4 described and summarised the steps and processes undertaken to conduct the exploratory study, and also presented the results and discussion. This chapter accounts for the Wireshark experiments with 20 participants in Cambodia. The chapter is organised as follows. Section 5.1 proposes the intent and design of this stage of the research while Section 5.2 gives a brief introduction to research participants from Cambodia. The ethics considerations and the recruitment of participants are detailed in Section 5.4 and Section 5.5, respectively. Steps and processes in conducting the workshop as well as the experiments are specified in Section 5.6, while the outcomes of the analysis are presented in Section 5.7. Section 5.8 discusses the outcomes and Section 5.9 concludes the findings.

5.1 Study intent and design

Two types of data were collected from Cambodian research participants. First, the researcher collected and analysed the HTTP conversation resulting from participants’ online browsing activities, and then conducted qualitative interviews with them. This chapter focuses only on the HTTP data. As stated in Chapter 3, this chapter aims to address the following research questions:

- What is the nature of information sharing in the HTTP headers?
  - 1. What type of information is being shared?
  - 2. With whom is this information being shared?
  - 3. What are the implications of this sharing?
As outlined in Section 4.5, SNSs are able to acquire the researcher’s intentionally and unintentionally shared information. This acquisition of information or tracking encouraged the researcher to find out about the perception and awareness of online users from Cambodia while also showing them the outcome of their online browsing experiments during qualitative interviews. The exploratory study in Chapter 4 confirmed that Wireshark is an appropriate tool to be used with participants. However, unencrypted passwords may be revealed in plain text format in Wireshark files, which needed to be considered when assessing risks to participants during the experiments. Before HTTP messages were collected from participants, a workshop was conducted by the researcher. The workshop aimed to introduce participants to the research intent, Wireshark application, basic HTTP protocols and data collection procedures. Participants took part in qualitative interviews after the researcher analysed the HTTP messages resulting from their online browsing.

5.2 Cambodian participants

Chapter 2 provides general background information of Cambodia, the country from which the researcher and the research participants originate. Meanwhile, this section briefly introduces the participants and the institution from which they were recruited. This research targeted Internet users, and the data collection, both experiments and interviews, was conducted in English. In order to maximise Cambodian participants’ understanding of the project details and procedures, interview questions, and consent forms, the researcher recruited the participants from the English department of the Institute of Foreign Language (IFL) at Royal University of Phnom Penh (RUPP).

RUPP is Cambodia’s oldest and largest public university, and was first established in January 1960 (RUPP, 2013a). It offers specialists in many fields including sciences, humanities, social sciences, and professional degrees in information technology, electronics, tourism, social work, and psychology (RUPP, 2013a). Language programs including English, French, Chinese, Japanese and Korean are offered through IFL (RUPP, 2013b).

The English language department at IFL offers a Bachelor of Education (BEd) in Teaching English as a Foreign Languages (TEFL) and the Bachelor of Arts (BA) in English for Work Skills. The BEd develops students’ skills in teaching English at secondary and tertiary levels whereas the BA enhances students’ ability to English language skills for a variety of employment where English proficiency is required. Since the course for both fields, BEd and BA, is conducted in English, the students from this department are known for possessing good English language skills. They also have access to computers and Internet facilities provided by
the university.

5.3 Sampling

Playing a major role in any research, a method of sampling is used if it is not possible to include the entire research population in the study. Sampling is a process of choosing a sample, a set of actual data sources, from a large set of possibilities or population of potential data sources in order to carry out empirical research (Pickard, 2007; Given, 2008). It is a two-step process applicable to both quantitative and qualitative research, and which includes (a) defining the population or the full set of possible data sources, and (b) selecting the actual sample (Given, 2008).

The number of the sample depends on a study’s purpose. In a qualitative study, the number of research participants is relatively smaller than that in a quantitative study (Kvale, 1996). More details on research paradigms will be provided in Chapter 7. This is because the qualitative approach does not intend to make inferences about the whole population based on a discovery as with the quantitative approach, but instead develops a rich understanding of a central phenomenon with relatively small samples or even a single case (Patton, 2002). Thus, sampling can be selected intentionally or purposefully based on what can be learned or understood from the phenomenon or context (Pickard, 2007). This strategy is known as purposive sampling and is used among qualitative researchers where they focus in-depth on small samples (Patton, 2002; Creswell, 2012). Purposive sampling requires the researcher to think critically and carefully about the criteria or parameters of the population being studied (Silverman, 2005); therefore it requires the population to be defined before selecting the actual sample (Given, 2008).

The power of purposeful sampling lies in selecting information-rich cases which provide insights and a deep understanding of the central phenomenon rather than generalisations; thus enabling the researcher to learn a great deal about the enquiry (Patton, 2002). Patton (2002, p. 244) also proposed that “There are no rules for sample size in qualitative inquiry.” For example, if the study’s purpose is to understand a phenomena as experienced by one specific person, then one participant is sufficient. In addition, to avoid drowning in more data than it is possible to analyse, small sample sizes are suitable for qualitative studies since they aim to emphasise an in-depth contextualised understanding of a specific phenomenon (Kvale, 1996; Given, 2008).

This research employed purposive sampling where the researcher recruited 20 participants for both stages of data collection: Wireshark experiments and qualitative interviews. The researcher spent three weeks in Cambodia to organise and conduct the workshop, the experi-
ments, and the interviews. Due to time and location constraints as well as the scope of research, it was considered that 20 participants would give the researcher quality time to prepare and arrange both experiments and interviews. In addition, as the focus of the research is online behaviour, and perception and awareness of online privacy among SNS users, the potential participants for the study would be participants who have regular access to the Internet and who also use SNS services. Only the participants who meet the following criteria were included in the study:

- Be of Cambodian nationality
- Be a member of academic staff or enrolled as a student at the Institute of Foreign Language (IFL), English department.
- Have at least one SNS account or profile such as Facebook, Twitter, or LinkedIn.
- Use the Internet on a regular basis (e.g., for checking e-mails, browsing online...).

5.4 Ethics consideration

As this study involved human participants, there were a number of ethical concerns to take into consideration including the language used in the interaction, consent, confidentiality, risk to the research participants, and data storage and accessibility.

Wireshark was chosen as a tool to collect the data, however, unencrypted passwords may be revealed in Wireshark files. In order to reduce the risk of participants’ passwords being revealed, the researcher informed participants about this possibility and also asked them to change the passwords of any associated account before the experiments began. The researcher also suggested that their new password contain the characters “liscsu” so that the researcher can identify and point out to them the unencrypted password of their visited sites.

The experiments and interviews were conducted in English, however, as the primary intended participants in this study were those whose primary Language is Other Than English (LOTE), the researcher recruited participants from IFL as they are known for possessing good English language skills. The researcher also helped the participants with direct translation if they had difficulties in language use, or in understanding the consent forms, Wireshark workshop, and interview questions. Ethics approval granted by Charles Sturt university (see Appendix A) was used to seek permission from IFL before the research details (see Appendix B) were advertised on the IFL bulletin board and before the recruitment of participants began.

The project sought volunteer participants only. Prior to Wireshark workshop, details of the project were clearly explained to participants, including how their information would be col-
lected and safeguarded, and then they were required to sign consent forms. Throughout this thesis, participants’ real names were replaced by a pseudonym or nickname. All personal information including the Wireshark files from their online browsing and conversations during the interviews, are safeguarded with password protected folders and computers, and are accessible only by the researcher.

Participants were reassured that if they became uncomfortable at any stage, they were permitted to withdraw from the study, and any identifiable information would be destroyed. Research data and confidential records, in both hard and soft copy, will be securely stored for five years after which time it will be destroyed.

5.5 Recruitment

The researcher recruited 20 participants who met the criteria outlined in Section 5.3. All subjects were required to participate in the workshop and in the two stages of data collection: Wireshark experiments and qualitative interviews. In order to gain access to potential participants, there are a number of possible means including working with the gatekeeper or an insider recruitment assistant, and advertising for participants via posters or distributing messages to an intended group (King & Horrocks, 2010). A gatekeeper can be someone who has the ability to grant or deny access to the participants, and an insider assistant can be someone who can distribute the project information details to the participants.

Soon after ethics approval was obtained from Charles Sturt University (see Appendix A), the researcher received help in advertising the project details at IFL’s information board for the recruitment from her sisters in Phnom Penh, before the researcher arrived in Cambodia. Information sheet (see Appendix B) detailed all the information about the Wireshark workshop, the experiments and the interview process. Interested participants who met the research criteria were encouraged to contact the researcher to register their interest.

5.6 Data collection and analysis

5.6.1 Wireshark workshop

A workshop was conducted by the researcher prior to Wireshark data collection. It took place in a conference venue at Lotus Blanc1 restaurant. The catering was prepared by students from

---

1http://www.yourphnompenh.com/lotus-blanc/
Pour un sourer d’enfant vocational training centre that supports poor or orphaned Cambodian children who cannot afford education.

The researcher explained the research details to participants in clear English. This research required that all participants perform online browsing while having their online activities recorded, saved, and finally collected by the researcher. It also required that they participate in the qualitative structured interviews where they would be interviewed by the researcher and where the outcomes of their online browsing activities would be revealed.

During the workshop, participants were introduced to basic information about HTTP cookies, information exchange in the HTTP headers, and the Wireshark application. As mentioned in Section 5.4, participants were advised about the potential risks involved in using Wireshark including the risk of revealing their passwords in Wireshark files that would be collected and analysed by the researcher. Therefore, the researcher suggested that participants change the passwords of their associated accounts to a new one containing the characters “liscsu” before the experiments began. They were also advised to change their passwords immediately after the experiments.

The researcher also assisted participants in installing and setting up Wireshark on their personal computers. Participants were also given a handout on research procedures (see details in Section 5.6.2), with the researcher providing a clear explanation of the procedures and consent forms (see Appendix C). From time to time, the researcher clarified that the procedures were well understood. She also used Khmer whenever any participant needed a clarification. Participants were then asked to sign the consent forms (see Appendix C) and the researcher reminded them that they could withdraw from the research at any stage if they felt uncomfortable.

### 5.6.2 Data collection

The Wireshark experiments involved each participant performing online browsing activities on their own computers while having these activities recorded by Wireshark. Unlike in the exploratory study where researcher selected sites and online activities based on the literature, the selection in this study was made by the participants. Each participant went through the same experimental procedures (specified in the handouts distributed by the researcher) detailed below:

- Making notes of what browsing activities will be performed.
- Terminating other applications running on the device which may be also using the HTTP protocol.
• Opening the browser and clearing all the cookies and search histories.
• Running Wireshark and starting to record the HTTP messages.
• Performing a set of actions as planned.
• Taking notes of what exact actions are performed if there are changes.
• Stopping the recording and saving the trace when the browsing actions are completed.

Again, the researcher periodically checked to see if participants had any questions regarding the procedures or the use of the Wireshark application. Once everything was set up, participants started browsing online and recording their activities. Then, they saved their Wireshark files in their full names. The researcher collected each file from each participant’s computer by transferring the file to her external hard drive for later analysis. For the purpose of participant privacy and confidentiality, each was assigned a nickname that will be used throughout this thesis, as shown in Appendix E.

5.6.3 Data analysis

Participants’ browsing activities can be grouped into four categories such as SNS usage, e-mail usage, online news reading, and online search as shown in Table 5.1. Unlike in the exploratory study, participants’ browsing did not include online shopping because this service is not yet available in Cambodia. However, participants visited sites like khmer79.com, cam11.com and bongthom.com and browsed gadgets such as mobile phones. While these sites do not provide any online shopping service, they instead serve as Web portals that bring together information from a variety of sources or Web sites such as online news, real estate agents, and gadget store Web sites. Although participants cannot purchase any item directly from those sites, they can browse for product details, in-store price and where to buy those products in Cambodia.

<table>
<thead>
<tr>
<th>SNSs</th>
<th>E-mail</th>
<th>News</th>
<th>Searches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>Gmail</td>
<td>khmer.rfi.fr</td>
<td>Google searches</td>
</tr>
<tr>
<td>Youtube</td>
<td>Yahoo</td>
<td>cnn.com</td>
<td>lookingtoday.com</td>
</tr>
<tr>
<td>Twitter</td>
<td></td>
<td>dap-news.com</td>
<td>Apple.com</td>
</tr>
<tr>
<td>Blogspot</td>
<td></td>
<td>plnompennpost.com</td>
<td>naruget.com</td>
</tr>
<tr>
<td>wordpress.com</td>
<td></td>
<td>sabay.com.kh</td>
<td>ohio.edu.au</td>
</tr>
<tr>
<td></td>
<td></td>
<td>postkhmer.com</td>
<td>wikipedia.com</td>
</tr>
<tr>
<td></td>
<td></td>
<td>bbc.com</td>
<td>eastasiaforum.org</td>
</tr>
<tr>
<td></td>
<td></td>
<td>khmer24.com</td>
<td>khmer79.com</td>
</tr>
<tr>
<td></td>
<td></td>
<td>cam111.com</td>
<td>cam111.com</td>
</tr>
<tr>
<td></td>
<td></td>
<td>cam111.com</td>
<td>bongthom.com</td>
</tr>
</tbody>
</table>

Table 5.1: Participants’ online activities and sites

52
The researcher examined the HTTP headers in each participant’s Wireshark file with the aim of identifying the types of shared information and third party sites. For example, let us examine HTTP headers resulting from Seila’s browsing at narutoget.com in Table 5.2 to see how the leaked information and third party sites were identified.

- (a) Seila’s browser sends a request to narutoget.com to retrieve the content.
- (b) narutoget.com returns an HTTP response which contains the page content (in HTML format) as well as Javascript code. The browser executes the code that it requires to fetch an advertisement banner from fastclick.net.
- (c) The browser then connects to fastclick.net server for the advertisement content. Within this request, narutoget.com also leaks Seila’s currently viewed video to fastclick.net.

<table>
<thead>
<tr>
<th>(a)</th>
<th>Host: narutoget.com\r\nConnection: keep-alive\r\nUser-Agent: Mozilla/5.0 (Windows NT 6.1) AppleWebKit/536.11 (KHTML, like Gecko) Chrome/20.0.1132.57 Safari/536.11\r\n</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b)</td>
<td>Location: <a href="http://www.narutoget.com%5Cr%5CnContent-Length">http://www.narutoget.com\r\nContent-Length</a>: 0\r\nSet-cookie: narutoget-cookies\nServer: lighttpd/1.4.28\r\n...\n&lt;script type=&quot;text/javascript&quot;&gt;</td>
</tr>
<tr>
<td>(c)</td>
<td>GET /w/get.media?sid=49560&amp;m=7&amp;tp=9&amp;t=d HTTP/1.1\r\nHost: media.fastclick.net\r\nConnection: keep-alive\r\nUser-Agent: Mozilla/5.0 (Windows NT 6.1) AppleWebKit/536.11 (KHTML, like Gecko) Chrome/20.0.1132.57 Safari/536.11\r\nReferer: <a href="http://www.narutoget.com/watch/episode">http://www.narutoget.com/watch/episode</a> details-english-subbed/\r\n</td>
</tr>
</tbody>
</table>

Table 5.2: Seila’s browsing at narutoget.com

In this case, narutoget.com - the first party site visited by Seila - was seen to leak Seila's currently viewed video details (episode number, title, subtitle details) to fastclick.net which is a third party site and which is also an advertising network.

5.7 Findings

Some findings of this study are consistent with those in the exploratory study. Participants’ visits to first party sites resulted in thousands of connections to various third party sites. Their information propagated from first party sites to third party sites, and from those third party sites to various other third party sites. Third party sites were identified as advertisers or data aggregators, as well as SNSs. The researcher classified the findings based on the level of
propagation, and the types of third party sites, while also pointing out the types of information being leaked or shared in the HTTP headers. Toolbars installed on some participants’ browsers were seen to cause their browser to contact third party site servers. However, no personally identifiable information was seen to be transferred to those servers. In addition, the sites visited by participants were not observed to leak any password information to third party sites.

5.7.1 First level traverse: from first party sites to third party sites

Third party sites which are advertisers or data aggregators

The most common third party sites in this study are advertisers for the first party sites or are purely data aggregators where advertisements are not visible on pages visited. Rany’s, Phary’s, Phara’s, and Rithy’s visits to an online news website khmer.rfi.com and dap-news.com resulted in numerous third party site connections as summarised respectively in Column (a) and Column (b) of Table 5.3. Kosal’s visits to his blog in wordpress.com also resulted in third party site connections as shown in Column (c) of Table 5.3.

<table>
<thead>
<tr>
<th>(a) Khmer.rfi.com</th>
<th>(b) dap-news.com</th>
<th>(c) wordpress.com</th>
</tr>
</thead>
<tbody>
<tr>
<td>googlefungation</td>
<td>bigman cambodia.com</td>
<td>s2.wp.com</td>
</tr>
<tr>
<td>horyzon-media.com</td>
<td>widdit.com</td>
<td>gravatar.com</td>
</tr>
<tr>
<td>siteestate.com</td>
<td>vigilink.com</td>
<td>Facebook</td>
</tr>
<tr>
<td>b.scorecardresearch.com</td>
<td>111.92.240.170</td>
<td>flickr.com</td>
</tr>
<tr>
<td>doubleclick.net</td>
<td>openx.sabay.com.kh</td>
<td>b.scorecardresearch.com</td>
</tr>
<tr>
<td>widdit.com</td>
<td>google.analytics.com</td>
<td>Twitter</td>
</tr>
<tr>
<td>googleapis.com</td>
<td>cdc.com.kh</td>
<td>quantserve.com</td>
</tr>
<tr>
<td>facebook.com</td>
<td>facebook.com</td>
<td>googleservices.com</td>
</tr>
<tr>
<td>twitter.com</td>
<td>googleanalytics.com</td>
<td>googleanalytics.com</td>
</tr>
<tr>
<td>vigilink.com</td>
<td>doubleclick.net</td>
<td>facebook.com</td>
</tr>
<tr>
<td>predicad.com</td>
<td>googleapis.com</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.3: List of third party sites

Meanwhile, Table 5.4 shows the way in which nytimes.com and phnompenhpost.com not only forwarded Seyha’s and Phara’s browser connections to third party sites, but also shared their clickstream or activities within the sites to those third party sites. Kosal’s visit to gsmarena.com resulted in his currently viewed page being shared among third party sites like b.scorecardresearch.com and google-analytics.com, as shown in Table 5.5.

During this study, 20/20 participants browsed Facebook. While 70% of the participants’ browsers were not seen to contact any third party servers while visiting Facebook, the browser of the other 30% (6/20 participants) contacted numerous third party sites, as illustrated in Table 5.6.

54
For instance, Rany used Dealply application on Facebook and Facebook forwarded the connections to various third party sites while also sharing Rany’s Facebook ID to them. Meanwhile, Makara clicked on an online news link posted by his friend on his Facebook page, sabay.com.kh while Seyha clicked on random advertisement. In addition, Mony used the Farmville application and Phanna used the Trove application. These activities caused connections to go to numerous servers outside Facebook.
Third party sites which are SNSs

Participants’ visited sites were seen to connect to various third party sites which are also SNSs such as Facebook, Twitter, and Google Plus. Those connections were necessary to fetch the contents of SNS widgets (e.g., Facebook’s Like button, Twitter’s Tweet button, and Google’s +1 button) to display on the first party site pages. For example, Phnompenhpost shared Phara’s currently visited page to Google while retrieving Google’s +1 button, as shown in Table 5.7.

Table 5.7: Phnompenhpost shared Phara’s visited page to Google

Among the sites visited by the participants in this study, excluding SNSs and e-mails, 79% of them (15/19 websites) connected to at least one SNS as a third party site, as shown in Table 5.8. The SNS that most commonly acted as a third party site is Facebook. Rather than simply relying on the Wireshark files alone, the researcher also checked each of these websites to find out if widgets are present on those first party site pages. Organisations like Apple, Wikipedia and Ohio University did not embed SNS widgets on their pages. Meanwhile, SNS widgets were not present on the BBC homepage, but Facebook and Twitter widgets are present on each article’s page.

<table>
<thead>
<tr>
<th>Visited sites</th>
<th>SNSs</th>
<th>Visited sites</th>
<th>SNSs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple.com</td>
<td>None</td>
<td>lookingtoday.com</td>
<td>F</td>
</tr>
<tr>
<td>bbc.com</td>
<td>F, T, none on the homepage</td>
<td>naruget.com</td>
<td>F, T</td>
</tr>
<tr>
<td>blogspot.com</td>
<td>F, T, G</td>
<td>nytimes.com</td>
<td>F, T</td>
</tr>
<tr>
<td>cnn.com</td>
<td>F, T</td>
<td>ohio.edu</td>
<td>None</td>
</tr>
<tr>
<td>eastasiaforum.org</td>
<td>F</td>
<td>phnompenhpost.com</td>
<td>F, T, G</td>
</tr>
<tr>
<td>gsmarena.com</td>
<td>F, T, G</td>
<td>postkhmer.com</td>
<td>F</td>
</tr>
<tr>
<td>Khmer.rf.fr</td>
<td>F, T, G</td>
<td>Wikipedia.com</td>
<td>None</td>
</tr>
<tr>
<td>khmer24.com</td>
<td>F</td>
<td>wordpress.com</td>
<td>F, T, G</td>
</tr>
<tr>
<td>khmer79.com</td>
<td>F</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.8: Visited sites and SNSs (F = Facebook, T = Twitter, G = Google)
5.7.2 Second level traverse: from third party sites to other third party sites

Participants’ browser connections propagated beyond one level. The three cases below help illustrate this finding.

Case 1: Seila’s browsing

While visiting narutoget.com, Seila’s browser connection also propagated to different levels of third party sites. The title of Seila’s currently viewed video propagated from (1st) narutoget.com to numerous third party sites including (3rd) rubiconproject.com, and from (3rd) rubiconproject.com to other third party sites including (4th) w55c.net. This (4th) w55c.net was seen to further connect to other third party sites such as (5th) adbrige.com, (5th) adadvisor.com and (5th) Facebook, as illustrated in Table 5.9. naruget.com was seen to connect to a long list of third party sites, as shown in Table 5.10 and the propagation is illustrated in Figure 5.1.

<table>
<thead>
<tr>
<th>(1st)</th>
<th>GET / HTTP/1.1\r\n Host: narutoget.com\r\n User-Agent: Mozilla/5.0 (Windows NT 6.1) AppleWebKit/536.11 (KHTML, like Gecko) Chrome/20.0.1132.57 Safari/536.11\r\n</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3rd)</td>
<td>GET /ad/8260.js HTTP/1.1\r\n Host: ads.rubiconproject.com\r\n User-Agent: Mozilla/5.0 (Windows NT 6.1) AppleWebKit/536.11 (KHTML, like Gecko) Chrome/20.0.1132.57 Safari/536.11\r\n Referer: <a href="http://www.narutoget.com/ads/728x90.html%5Cr%5Cn">http://www.narutoget.com/ads/728x90.html\r\n</a></td>
</tr>
<tr>
<td>(4th)</td>
<td>GET /ct/rubicon-cms2.html HTTP/1.1\r\n Host: cti.w55c.net\r\n User-Agent: Mozilla/5.0 (Windows NT 6.1) AppleWebKit/536.11 (KHTML, like Gecko) Chrome/20.0.1132.57 Safari/536.11\r\n Referer: <a href="http://tap2-cdn.rubiconproject.com/Seila%E2%80%99s">http://tap2-cdn.rubiconproject.com/Seila’s</a> country location</td>
</tr>
<tr>
<td>(5th)</td>
<td>GET /fr/u.php?p=207350426053595&amp;m= ... HTTP/1.1\r\n Host: <a href="http://www.facebook.com%5Cr%5Cn">www.facebook.com\r\n</a> User-Agent: Mozilla/5.0 (Windows NT 6.1) AppleWebKit/536.11 (KHTML, like Gecko) Chrome/20.0.1132.57 Safari/536.11\r\n Referer: <a href="http://cti.w55c.net/ct/rubicon-cms2.html%5Cr%5Cn">http://cti.w55c.net/ct/rubicon-cms2.html\r\n</a></td>
</tr>
</tbody>
</table>

Table 5.9: HTTP headers: Seila’s browsing propagation

Case 2: Makara’s browsing

Makara’s searches in the Google search engine were seen to be transferred from (1st) Google to (2nd) twilight-hikari.deviantar.com (one of the search results provided by Google search en-
(1st) narutoget.com
fastclick.net
crunchyroll.com
hulium.com
apmebf.com
doubleclick.net
google-analytics.com
game-advertising.online.com
adnxs.com
bluekai.com
admeld.com
rubiconproject.com
whois.amung.us
extreme-dn.com
uac.advertising.com
narutochatt.chatango.com
yieldmanager.com
amp.yahoo.com
b.scorecardresearch.com
content.yieldmanager.eggsuite.net
adsmarket.com
Facebook
Twitter
tracking.traviangames.com
browsegame.travian.com

(3rd) js
script.gx101.com
quantserve.com
rhub.com
yieldmanager.com
rtbidder.net
w55c.net
bluekai.com
turn.com
gwallet.com
everesttcc.ch
pulsesmgr.com
doubleclick.net
rubicon-match.dotomi.com
advertising.com
netseer.com
mathtag.com
owneriq.net
simpli.fi
media6degrees.com
lucidmedia.com
ru4.com
veruta.com
triggit.com
mookiel.com
adserver.org
contextweb.com
pubmatic.com
invitemedia.com
rightaction.com
openx.net
agkn.com

(4th) js
adbrite.com
adap.tv
Facebook
p-td.com
adadvisor.net
tinyurl.com

Table 5.10: Seila’s browsing propagation

Figure 5.1: narutoget.com traverse map

gine), then from (2nd) twilight-hikari.deviantar.com to numerous other third party sites including (3rd) da-ads.com. da-ads.com was also seen to forward the keywords to other third party
sites such as (4th) googleadservices.com, (4th) doubleclick.net and (4th) rubiconproject.com. After this, the connection further connected to numerous other third party sites including (5th) yieldmanager.com, as shown in Table 5.11.

<table>
<thead>
<tr>
<th>(1st) Google</th>
<th>(2nd) twilight-hikari.deviantar</th>
<th>(3rd) da-ads.com</th>
<th>(4th) rubiconproject.com</th>
</tr>
</thead>
</table>

Table 5.11: Makara’s information propagation

Case 3: Phanna

Phanna first visited his Facebook profile, then used or clicked (3rd) Trove (a third party application on Facebook). (3rd) Trove then was seen to connect to other third party sites including (4th) YouTube. (4th) Youtube then connected to (5th) yimg.com which then further connected to (6th) eBay.com, as shown in Table 5.12.

<table>
<thead>
<tr>
<th>(1st) Facebook</th>
<th>(3rd) Trove</th>
<th>(4th) Youtube</th>
<th>(5th) yimg.com</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3rd) fb.trove.com</td>
<td>(4th) Youtube.com, google-analytics.com</td>
<td>(5th) yimg.com</td>
<td>(6th) eBay.com</td>
</tr>
</tbody>
</table>

Table 5.12: Phanna’s information propagation

5.7.3 Browser toolbar connections

In addition to extra connections to third party sites caused by first party sites, participants’ browser toolbars also caused extra connections to third party sites. Nary’s and Rithy’s browser toolbars sent out connections to third party sites including Google and Yahoo in the two cases below. However, these connections did not transfer any of the participants’ identifiable information to third party sites.

59
Case 1: Nary and the *Babylon* toolbar

Nary’s browser connected to the (a) *montiera.com* server in order to retrieve the *Babylon* toolbar and to establish the connection. (b) *montiera.com* then replied with a script necessary to establish the connection, while also setting cookies to Nary’s computer, as shown in Table 5.13. Her browser executed the script that resulted in another connection to (c) *babylon.com*. Then (d) *babylon.com* set cookies on Nary’s computer. *Babylon* connected to Google and its services: (a) *google-analytics*, (b) *googleadservices* and (c) *doubleclick.net* while also sharing the cookie values to those third party sites, as shown in Table 5.14.

![Request Example](https://example.com/request)

**Table 5.13: Nary’s toolbar connection establishment**

Case 2: Rithy and the *Incredibar* toolbar

Similarly to Nary’s case, Rithy’s browser first connected to the (a) *Incredibar* server because of the *Incredibar* toolbar installed on his browser. *Incredibar* was then seen to connect to various third party sites including (b) Yahoo and (c) Google, as shown in Table 5.15. Nary’s and Rithy’s toolbars’ connections to third party sites are summarised in Table 5.16.

![Request Example](https://example.com/request)
5.8 Discussion

5.8.1 Tracking by third party sites which are advertisers or data aggregators

As usual, Moly checked her Yahoo mail and used her Facebook account. At the same time, she also read the news on CNN while also logging into her account on Lookbook\(^2\) where

\(^2\)http://lookbook.nu/
she accessed a collection of fashion pictures such as models and clothing lines. Lookbook, Yahoo, and CNN were seen to forward her activities within those sites to numerous third party sites including Scorecard Research as shown in Table 5.17. Therefore, Scorecard Research can track her activities across different pages on Lookbook, Yahoo, and CNN, as shown in Table 5.18. In addition, Nary’s and Rithy’s browser toolbars also caused extra connections from their browsers to third party sites like Google and Yahoo. Some common third party sites which act as advertisers for first party sites are Scorecard Research, Google, and Yahoo. Therefore, these advertisers are able to track participants’ online movement across different first party sites.

<table>
<thead>
<tr>
<th>lookbook.com</th>
<th>Yahoo mail</th>
<th>cnn.com</th>
<th>Facebook</th>
</tr>
</thead>
<tbody>
<tr>
<td>quantserve.com</td>
<td>yimg.com</td>
<td>turner.com</td>
<td>Facebook CDN</td>
</tr>
<tr>
<td>b.scorecardresearch.com</td>
<td>b.scorecardresearch.com</td>
<td>b.scorecardresearch.com</td>
<td></td>
</tr>
<tr>
<td>googleapis.com</td>
<td>yieldmanager.com</td>
<td>imrworldwide.com</td>
<td></td>
</tr>
<tr>
<td>Facebook</td>
<td>bs.serving-sys.com</td>
<td>Facebook</td>
<td></td>
</tr>
<tr>
<td>Twitter</td>
<td>ib.adnxs.com</td>
<td>Twitter</td>
<td></td>
</tr>
<tr>
<td>google-analytics.com</td>
<td>avatars.zenfs.com</td>
<td>chartbeat.com</td>
<td></td>
</tr>
<tr>
<td>sponsorads.de</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lbstatic.nu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>googleservices.com</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>doubleclick.net</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>google syndication.com</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>adsbysocket.com</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mediaplex.com</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>apmefb.com</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>amazonaws.com</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pointroll.com</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shopnastygal.com</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>file.storetsen.cafe24.com</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rackcdn.com</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>new relic.com</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.17: Moly’s browsing activities and associated third party sites

<table>
<thead>
<tr>
<th>Advertiser</th>
<th>Moly’s browsing activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scorecard Research</td>
<td>lookbook.com, Yahoo mail, cnn.com</td>
</tr>
</tbody>
</table>

Table 5.18: Moly’s visited pages able to be tracked by Scorecard Research

### 5.8.2 Tracking by third party sites which are SNSs

Within one browsing session, Kosal checked his Facebook and Twitter social network profiles while also logging into his blogs at Wordpress and Blogspot. He used Gmail for his e-mail service. He also checked mobile phone prices and trends on Khmer24, GSM Arena and Apple sites. Based on his browsing activities, and the third party sites detected in the Wireshark file,
it is observed that Web sites like Wordpress, Blogspot, Khmer24, and GSM Arena transferred information about Kosal’s movements within those sites to numerous third party sites as shown in Table 5.19.

<table>
<thead>
<tr>
<th>(a) Facebook</th>
<th>(b) Twitter</th>
<th>(c) Gmail</th>
<th>(d) Apple</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook CDN</td>
<td>Twitter CDN</td>
<td>Encrypted</td>
<td>Apple CDN</td>
</tr>
<tr>
<td>google-analytics.com</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e) wordpress.com</td>
<td>(f) blogspot.com</td>
<td>(g) khmer24.com</td>
<td>(f) gsmarena.com</td>
</tr>
<tr>
<td>Facebook</td>
<td>Twitter</td>
<td>bigmancambodia.com</td>
<td>Facebook</td>
</tr>
<tr>
<td>googleapis.com</td>
<td>googleanalytics.com</td>
<td>google-analytics.com</td>
<td>Twitter</td>
</tr>
<tr>
<td>s2.wp.com</td>
<td></td>
<td></td>
<td>googleapis.com</td>
</tr>
<tr>
<td>gravatar.com</td>
<td></td>
<td></td>
<td>adn.ebay.com</td>
</tr>
<tr>
<td>flickr.com</td>
<td></td>
<td></td>
<td>google-analytics.com</td>
</tr>
<tr>
<td>b.scorecardresearch.com</td>
<td></td>
<td></td>
<td>exponential.com</td>
</tr>
<tr>
<td>quantserve.com</td>
<td>googleapis.com</td>
<td></td>
<td>tribalfusion.com</td>
</tr>
<tr>
<td>googleanalytics.com</td>
<td></td>
<td></td>
<td>mediaplex.com</td>
</tr>
<tr>
<td>googleservices.com</td>
<td></td>
<td></td>
<td>fnmpub.net</td>
</tr>
<tr>
<td>google-analytics.com</td>
<td></td>
<td></td>
<td>doubleclick.net</td>
</tr>
<tr>
<td>doubleclick.net</td>
<td></td>
<td></td>
<td>intellicx.com</td>
</tr>
<tr>
<td>googleservices.com</td>
<td></td>
<td></td>
<td>betrad.com</td>
</tr>
<tr>
<td>(e) wordpress.com</td>
<td>(f) blogspot.com</td>
<td>(g) khmer24.com</td>
<td>(f) gsmarena.com</td>
</tr>
<tr>
<td>Facebook</td>
<td>Twitter</td>
<td>bigmancambodia.com</td>
<td>Facebook</td>
</tr>
<tr>
<td>googleapis.com</td>
<td>googleanalytics.com</td>
<td>google-analytics.com</td>
<td>Twitter</td>
</tr>
<tr>
<td>s2.wp.com</td>
<td></td>
<td></td>
<td>googleapis.com</td>
</tr>
<tr>
<td>gravatar.com</td>
<td></td>
<td></td>
<td>adn.ebay.com</td>
</tr>
<tr>
<td>flickr.com</td>
<td></td>
<td></td>
<td>google-analytics.com</td>
</tr>
<tr>
<td>b.scorecardresearch.com</td>
<td></td>
<td></td>
<td>exponential.com</td>
</tr>
<tr>
<td>quantserve.com</td>
<td>googleapis.com</td>
<td></td>
<td>tribalfusion.com</td>
</tr>
<tr>
<td>googleanalytics.com</td>
<td></td>
<td></td>
<td>mediaplex.com</td>
</tr>
<tr>
<td>googleservices.com</td>
<td></td>
<td></td>
<td>fnmpub.net</td>
</tr>
<tr>
<td>google-analytics.com</td>
<td></td>
<td></td>
<td>doubleclick.net</td>
</tr>
<tr>
<td>doubleclick.net</td>
<td></td>
<td></td>
<td>intellicx.com</td>
</tr>
<tr>
<td>googleservices.com</td>
<td></td>
<td></td>
<td>betrad.com</td>
</tr>
<tr>
<td>(e) wordpress.com</td>
<td>(f) blogspot.com</td>
<td>(g) khmer24.com</td>
<td>(f) gsmarena.com</td>
</tr>
<tr>
<td>Facebook</td>
<td>Twitter</td>
<td>bigmancambodia.com</td>
<td>Facebook</td>
</tr>
<tr>
<td>googleapis.com</td>
<td>googleanalytics.com</td>
<td>google-analytics.com</td>
<td>Twitter</td>
</tr>
<tr>
<td>s2.wp.com</td>
<td></td>
<td></td>
<td>googleapis.com</td>
</tr>
<tr>
<td>gravatar.com</td>
<td></td>
<td></td>
<td>adn.ebay.com</td>
</tr>
<tr>
<td>flickr.com</td>
<td></td>
<td></td>
<td>google-analytics.com</td>
</tr>
<tr>
<td>b.scorecardresearch.com</td>
<td></td>
<td></td>
<td>exponential.com</td>
</tr>
<tr>
<td>quantserve.com</td>
<td>googleapis.com</td>
<td></td>
<td>tribalfusion.com</td>
</tr>
<tr>
<td>googleanalytics.com</td>
<td></td>
<td></td>
<td>mediaplex.com</td>
</tr>
<tr>
<td>googleservices.com</td>
<td></td>
<td></td>
<td>fnmpub.net</td>
</tr>
<tr>
<td>google-analytics.com</td>
<td></td>
<td></td>
<td>doubleclick.net</td>
</tr>
<tr>
<td>doubleclick.net</td>
<td></td>
<td></td>
<td>intellicx.com</td>
</tr>
<tr>
<td>googleservices.com</td>
<td></td>
<td></td>
<td>betrad.com</td>
</tr>
</tbody>
</table>

Table 5.19: Kosal’s browsing activities and associated third party sites

SNSs like Facebook, Twitter and Google Plus are observed to be among third party sites which track participants’ online movements across different sites due to the fact that visited sites had SNS widgets embedded on their pages (e.g., Facebook’s Like button, Twitter’s Tweet button, and Google’s +1 button). As a result, Facebook, Twitter, and Google are able to know Kosal’s movements on any sites on which their widgets are embedded, as summarised in Table 5.20, in addition to information voluntarily provided by Kosal to those SNSs. Even though Twitter and Khmer24 do not have Google button on their page, Google is still able to know Kosal’s activities because of the connection to google-analytics.com.

<table>
<thead>
<tr>
<th>SNSs</th>
<th>Kosal’s browsing activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>wordpress.com</td>
</tr>
<tr>
<td></td>
<td>blogspot.com</td>
</tr>
<tr>
<td></td>
<td>khmer24.com</td>
</tr>
<tr>
<td></td>
<td>gsmarena.com</td>
</tr>
<tr>
<td>Twitter</td>
<td>wordpress.com</td>
</tr>
<tr>
<td></td>
<td>blogspot.com</td>
</tr>
<tr>
<td></td>
<td>gsmarena.com</td>
</tr>
<tr>
<td>Google</td>
<td>Gmail account</td>
</tr>
<tr>
<td></td>
<td>wordpress.com</td>
</tr>
<tr>
<td></td>
<td>blogspot.com</td>
</tr>
<tr>
<td></td>
<td>khmer24.com</td>
</tr>
<tr>
<td></td>
<td>gsmarena.com</td>
</tr>
<tr>
<td></td>
<td>Twitter</td>
</tr>
</tbody>
</table>

Table 5.20: Kosal’s browsing activities able to be tracked by SNSs
5.9 Conclusion

Similarly to the researcher’s browsing in the exploratory study, participants’ visits to first party sites resulted in HTTP connections to numerous servers. Some connections are necessary to fetch the content from the first party site CDN while other connections are going to third party sites which were not directly requested by participants. Some HTTP connections resulting from the participants’ browsing propagated to more than just one level of third party sites. Information on participants’ browsing activities was seen to be shared or leaked from first party sites to third party sites, and from those third party sites to other third party sites. Their browser toolbars also created connections to third party sites such as Google. These connections were not detected in the exploratory study, and have not been mentioned in the existing literature. Third party sites in these experiments were identified as advertisers, data aggregators, and SNSs.

The sharing of participants’ information and browsing activities to third party sites enabled those third party sites to track participants’ movements. For example, within Kosal’s browsing session, sites like Wordpress, Blogspot, Khmer24, and GSM Arena shared Kosal’s browsing activities to Facebook while retrieving Facebook’s Like button. As a result, Facebook acquired Kosal’s movements among those first party sites within that browsing session, in addition to his information populated on his Facebook profile. Meanwhile, the toolbars were not seen to share any participants’ identifiable information to third party sites. However, the connection to the same third party sites, via the toolbars and the first party sites, eventually enabled those third party sites to link the activities with a specific participant.

Experiments conducted with 20 participants allowed the researcher to see another aspect of information sharing via the HTTP header. For example, neither the exploratory study, nor the existing literature has detected the sharing of online users’ information from their browser toolbars to third party sites. In addition, 19/20 participants did not use any browser extension to prevent online tracking. Therefore, the number of third party sites contacted in these experiments appear to be higher than in the case of the researcher’s browsing with AdBlock Plus extension. For example, Seila’s visit to narutoget.com caused his information to propagate to a long list of third party sites, and from those third party sites to another long list of third party sites, as summarised in Table 5.9. With or without the protection, connections to third party sites still exist.

The findings in the exploratory study as well as in this chapter helped the researcher to develop and limit the scope of the interview questions. The next stage of this thesis focuses on participants’ opinions and awareness of online privacy and online tracking, particularly tracking by SNSs. This research proceeds to elucidate what privacy means to participants, and how they
react to the fact that their browsing activities are being monitored by various third party sites including Facebook, Twitter and Google Plus.
Chapter 6

Social network sites and privacy: the literature

As found in Chapter 4 and Chapter 5, Facebook, Twitter and Google Plus acquired the researcher’s and participants’ unintentionally shared information in addition to information voluntarily provided to these SNSs. This finding encouraged the researcher to explore participants’ perception and awareness regarding privacy and the practice of online tracking by SNSs, based on their browsing activities. This chapter examines the existing literature on SNSs and associated privacy concerns. First it presents the definition and components that constitute an SNS system in Section 6.1. Then in Section 6.2, it provides a background on users’ behaviour on SNSs, particularly on their interaction and the types of information they revealed on SNSs. Section 6.3 of the chapter presents a general understanding of the meaning and importance of privacy while Section 6.4 examines the literature on users’ awareness of online tracking by third party sites. Meanwhile, details on the methodology and technique employed in this stage of the research in terms of data collection and analysis are described in Chapter 7.

6.1 Social network sites

Social network sites (SNSs) such as MySpace, LinkedIn, Facebook, Twitter and Google Plus have attracted millions of users worldwide. There are many SNSs that support a wide range of interests, audiences, languages and cultures (boyd & Ellison, 2007). For example, while Facebook is not available in China, there are other SNSs such as Renren.com and Kaixin001.com that enable people in China to communicate with each other. As stated in Section 1.7, for the purpose of this research, SNSs refer to “web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other
users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system” (boyd & Ellison, 2007, p.211). SNSs break down barriers created by time, distance and culture among Internet users and create an online group or community where those users can interact with one another by sharing opinions, insights, information, interest and experiences (Reynolds, 2012).

According to boyd and Ellison (2007), the first recognisable SNS, Sixdegrees.com, was launched in 1997. Similarly to current SNSs, Sixdegrees.com allows users to create profiles, and search for and connect with their friends. Despite the fact that it attracted millions of users, Sixdegrees.com was not successful and terminated its service in 2000. Since 1997, various categories of SNSs emerged: personal, professional, and dating. Similar to LinkedIn.com, Ryze.com was launched in 2001 to help people leverage their business and professional networks. Other SNSs like Tribe.net and Friendster also emerged and became popular, before other SNSs such as MySpace, Last.FM, Twitter and Facebook also came along (as shown in Figure 6.1). Facebook is seen to have grown very quickly in terms of numbers of users since it first started in 2004 (Yadav, 2006), and has currently attracted more than one billion active users (Ong, 2013).

In order to join an SNS and sign up for an account, users are typically required to provide personal information such as name, gender, date of birth, location, photo and e-mail address. While many SNSs implemented various technological features and affordances, they usually allow their users to “type oneself into being” on profile pages (boyd & Ellison, 2007). Once users have an SNS account, they are able to connect with or “befriend” others by sending out or accepting friend requests via SNS profiles as well as e-mails. SNSs also provide a “search” feature where users can look up people they know who may already have an SNS account within the same network. Users also have the ability to remove or block a contact from their list of friends, or to keep their profile in a completely private mode where no one else apart from the profile owners, can view the posted information.

SNS users are usually encouraged to share as much information as SNS features can support to enhance this sharing experience. In addition to basic information required to sign up for an SNS account, SNS users are also encouraged to relate stories about themselves and their lives. For example, Facebook and Google Plus enable users to share information about their educational background, work experience, hometown, and interests. By default, an SNS profile is made available to the public or at least to other users within the same network (Krishnamurthy & Wills, 2008). However, users can change the default privacy setting in order to limit their audience. Some users may not be aware of the default settings while others do not have enough technical knowledge to change them - or they simply may not care about privacy settings (Krishnamurthy, Gill, & Arlitt, 2008). In this way, users may be revealing information about themselves while being unaware that this information is available to an unintended audience.
As mentioned in Chapter 3, in addition to providing a means for communication, some SNSs like MySpace, Hi5 and Facebook also allow their users to use third party applications such as games. Those applications are not hosted by SNSs, rather, they reside on third party servers. These make SNSs even more interactive because, for example, users have the ability to play games (e.g., FarmVille) with their SNS friends. SNSs are now also used for other purposes such as business and entertainment due to the fact that SNSs have attracted large numbers of users meaning advertisements or promotions via SNSs can reach many users very quickly. For

(Gross & Acquisti, 2005; boyd & Ellison, 2007).

Figure 6.1: Timeline of the launch dates of major SNSs
(boyd & Ellison, 2007, p.212)
example, Speedo ¹, a manufacturer and distributor of swimwear that was founded in Australia in 1914, uses Facebook to advertise their new products or promotions. A television music competition franchise like The X Factor also uses Twitter to keep its audience updated about the show and the contestants.

6.2 Social network sites and online users

6.2.1 The use of SNSs

A number of studies suggest that online users are usually active on SNSs in order to maintain ties and contacts with people they know in real life (e.g., connecting with people they recently met or keeping in touch with friends), express their feelings and share their thoughts, and use third party applications available on SNSs such as Facebook quizzes or Hi5 e-postcards (Joinson, 2008; Pempek, Yermolayeva, & Calvert, 2009; Young, 2009; Ross et al., 2009; Al-Saggaf, 2011).

The majority of the participants surveyed by Acquisti and Gross (2006) reported that they use Facebook mainly because Facebook allows them to learn about and find their classmates while also allowing other people to get in touch with them. They denied the idea that Facebook is useful for dating or self-promotion. Other participants also claim that SNSs are a convenient way for them to maintain their interpersonal interconnectivity and social enhancement, and that SNSs provide instant communication and connection with their friends (Cheung, Chiu, & Lee, 2011). In addition, SNS users, particularly those on Facebook, tend to search for people whom they know offline, rather than for complete strangers (Lampe, Ellison, & Steinfield, 2006).

SNSs, particularly Facebook, help students to maintain social capital among their friends (Ellison, Steinfield, & Lampe, 2007). Social capital refers to “the sum of the resources, actual or virtual, that accrue to an individual or a group by virtue of possessing a durable network of more or less institutionalised relationships of mutual acquaintance and recognition” (Bourdieu & Wacquant, 1992, p.14). Bridging social capital refers to weak ties that are connections between interactants who may provide useful information or perspective, but not emotional support. On the contrary, bonding social capital reflects stronger ties between individuals in a close relationship, who may provide emotional support (Putnam, 2000). For example, when students graduate and move away from college, Facebook helps to bring them and their peers together.

¹www.speedo.com.au
regardless of the physical space. Online interactions support relationships and help people keep in touch, but do not remove existing connections or relationships from the offline world.

6.2.2 SNS users and the disclosure of information

As mentioned previously, in order to be able to connect and communicate with others on SNSs, users are usually encouraged to share information about themselves. In addition to information required by SNSs at the sign-up stage, information generated and revealed by users while engaging with others on SNSs tends to be personal and related to their hobbies and interests (Gross & Acquisti, 2005; Taraszow, Aristodemou, Shitta, Laouris, & Arsoy, n.d.; Taddicken, 2014). Gross and Acquisti (2005) examined the amount of information disclosed on Facebook among 4,000 students at Carnegie Mellon university, and their usage of the sites’ privacy settings. Students’ profiles contain their profile pictures, and information about their date of birth, phone number, current location or residence, relationship status, political views and their hobbies (e.g., music, books and movies). Also, those who list their relationship status as other than single also identify and link their partner’s Facebook profile to theirs. Similarly in a more recent work, 1,200 students at Pace University, New York, were also reported to share their phone numbers and home addresses on top of other basic information revealed in their SNS profiles (Lawler & Molluzzo, 2010).

Multiple factors could influence self-disclosure online. In addition to the encouragement by SNSs, users’ sharing of information about themselves is also encouraged by their tendency to understand of self and identity that is constituted through interaction with others (Pullinger, 2001). Cyberspace provides a new opportunity to present a new image of ourselves. SNSs are seen to provide such an opportunity, particularly for young teenagers, to establish their online personal identity (Livingstone, 2008; Pempek et al., 2009). There is also a notion that interactants can develop personalised relationships characterised by intimacy (Walther, 1992). Computer-mediated communication (CMC) is seen to be limited compared to real life communication, thus, online users attempted to compensate by hyper-personalising themselves in order to gain confidence in their brief interaction and by exchanging more intimate conversation and disclosure with other interactants (Tidwell & Walther, 2002).

In addition, online users tended to disclose rich information about themselves on SNSs because of the convenience in sharing, the relationship building with other SNS users (Waters & Ackerman, 2011), and the enjoyment experienced while using SNS services (Krasnova et al., 2010). For example, on Facebook, users can connect not only with their offline friends, but also with other Facebook users whom they have never met. Users can also maintain or strengthen that connection by participating in any third party application service enabled by Facebook. And
it is only one-click-away from sharing information with their connections (Al-Saggaf, 2011; Krasnova et al., 2010). However, users who are concerned with privacy tend to share less about themselves on SNSs (Staddon, Huffaker, & Brown, 2012; Taddicken, 2014).

6.3 Privacy

As this stage of the research focuses on Cambodian SNS users’ perception and awareness of privacy and online tracking, this section examines the meaning and importance of privacy from the classical philosophical theories in the literature.

6.3.1 Definition

Privacy is becoming one of the most enduring social issues associated with information technologies. There have been many efforts to define privacy, yet there has never been an agreement on a single definition or theory of what privacy is. Privacy is often described in different terms and notions. It can be described as to be intruded upon, lost, violated, diminished, or breached. Each description reflects the conception and perception of privacy from different models or theories of privacy. For example, from a right-based perspective, privacy was once described as the right to be left alone which is an element of a person’s right to life (Warren & Brandeis, 1890). On the other hand, from an interest-based perspective, privacy can be defined as “the interest individuals have in sustaining a personal space, free from interference by other people and organisation” (Clarke, 1997). Meanwhile, the Oxford Dictionaries defines privacy as “A state in which one is not observed or disturbed by other people”, and “The state of being free from public attention”.

Privacy is also described in terms of notions of control and limitation. For example, Westin (1966, p.431) views privacy as “the claim of an individual to determine what information about himself or herself should be known to others... This also, involves when such information will be obtained and what uses will be made of it by others.” Fried (1990, p.54) relates privacy to the limiting or controlling knowledge of others about oneself - “the less that is known about us the more privacy we have.” He also adds that knowing general facts about a person is acceptable, but knowing details leads to invasion of that person’s privacy.

On the other hand, Gavison (1983, p.425-426) argues that the notion of privacy as control is misleading and ambiguous. She defines privacy as the limitation of others’ access to an individual and that limitation consists of three elements known as secrecy (when one has control

\[2\text{http://www.oxforddictionaries.com/definition/english/privacy}\]
over one’s information), anonymity (when there is no one paying attention to us), and solitude (when others have no physical access to us). In her view, “... in a perfect privacy, no one has any information about X, no one pays any attention to X, and no one has physical access to X” (Gavison, 1983, p.428). Hence, privacy is lost when others obtain information about us, pay attention to what we do, and/or have physical access to us. She also adds that privacy is not about all or nothing: perfect privacy is impossible, so is the total loss of privacy.

Meanwhile, Parent (1983) suggests that the definition of privacy should be consistent with ordinary language so that people can talk consistently, clearly and precisely about the family or concepts to which privacy belongs. He defines privacy as “the condition of not having undocumented personal knowledge about one possessed by others” (Parent, 1983, p.269). A person’s privacy is diminished when others possess knowledge about an individual. Achieving consistency in the definition of privacy could be challenging as different countries and cultures also have different perceptions of privacy and what it means. For example, Japan does not have a word for privacy; rather they adopted the English word and presented it in Katakana phonetic syllabary as a word of foreign origin (Mizutani, Dorsey, & Moor, 2004).

Kang (1998) clustered the term “Privacy” into three groupings that are functionally interconnected and simultaneously implicated by the same event or practice. The first cluster of privacy, one’s physical space, refers to one’s territorial solitude that is shielded or protected from invasion. The second cluster of privacy is concerned with choice - which is an individual’s ability to make decisions without interferences. Last but not least, the third cluster of privacy concerns the flow of personal information. This cluster concerns an individual’s control over the processing and the use of personal information.

6.3.2 Personal or private information

Section 6.3.1 examined some philosophical theories about privacy. This section will look at what is considered personal or private information. Personal information consists of facts about individuals that most would choose not to reveal to society (Parent, 1983). The term “personal” does not mean sensitive or embarrassing (Kang, 1998). According to Kang (1998), personal or private information describes the relationship between the information and an individual and to some extent, it is identifiable to that individual.

People can be sensitive to a different degree about those facts: some can be very sensitive in that they choose not to reveal them, other than to close friends and family, while others do not really care if those facts are widely known. Parent (1983) also added that some facts which were once considered personal could also change over time. In his example, people’s sexual
preferences were once very personal and most people did not reveal or talk about them, but today in some parts of the world, people can talk about this topic openly.

Al-Saggaf and Weckert (2011) pointed out that our inner thoughts and feelings, our personal relationships, our information (particularly about our lives, health and finance), our own space (e.g., our house, desk, room), and the state of being unobserved should be treated as an individual’s own business and should be private matters, at least in a sense that those involved are able to choose what details they intend to share and with whom. People might not mind others knowing various things about them, but the control over access to the knowledge is crucial (Al-Saggaf & Weckert, 2011).

Similarly, in Australia’s Privacy Act, personal information refers to “… information or an opinion (including information or an opinion forming part of a database), whether true or not, and whether recorded in a material form or not, about an individual whose identity is apparent, or can reasonably be ascertained, from the information or opinion” (OAIC, 2012). This information can be identifiable to a specific individual and may also include an individual’s name, address, medical records, bank account details, photos, videos, preferences, opinions and workplace.

6.3.3 Importance of privacy

Clarke (1997) classified the importance of privacy into four groups: psychological, sociological, economical and political. Psychologically speaking, privacy provides a private space to an individual where he or she is able to judge if anything or anyone in the vicinity is a threat. Sociologically, privacy also enables an individual to be free to behave and to associate with others in the society. Clarke (1997) also adds that freedom gives people space to innovate and be creative, therefore, economically allows the competition to be fierce. Last but not least, privacy also provides people with the freedom to think argue and act, and thus enable democracy, which can be threatened by surveillance.

Privacy is also essential to achieve important human ends such as trust, friendship and intimacy (Fried, 1990). In Fried’s view, the intimacy of relationships between individuals depends on the amount of information they disclose to each other. For example, we tend to share more personal information about ourselves to our close friends than to the public or a complete stranger. In addition to the sharing of information, privacy also enables us to form a variety of social relationships with other people depending on how we act and behave towards each other because there are different patterns of behaviour associated with different relationships (Rachels, 1975). For example, John may appear to be very playful and affectionate with his
children, while at work, he appears to be more formal with his boss and colleagues, and to his close friends, he may show aspects of his personality of which his children and his boss are not aware.

Privacy is also perceived as important or necessary in the sense that it protects an individual from embarrassment, discrimination, and unfairness (Rachels, 1975; Kang, 1998). For example, revealing a medical condition may result in a person losing their job or being rejected from an insurance application. Similarly to Clarke’s economical classification (1997), privacy also protects one’s benefits in a competitive situation. For instance, it would be a disadvantage to ABC Company if their business plan were known by their competitors. In addition, privacy protects us from scrutiny, prejudice, coercion and pressure to conform, and thus becomes valuable in terms of an individual’s freedom and independence, which are essential for democracy (Tavani, 2007a).

### 6.4 SNS users and privacy

#### 6.4.1 SNSs’ privacy settings and privacy policies

An issue raised earlier in this chapter is that, by default, the majority of SNSs’ privacy settings allow visibility of users’ profiles to strangers or other members of SNSs who are not necessarily connected to users (Krishnamurthy & Wills, 2008). Users do not change their default privacy settings mainly because they were not aware that they have the ability to do so (Acquisti & Gross, 2006) or they have difficulty in understanding the privacy settings (Lipford, Besmer, & Watson, n.d.).

The study by Gross and Acquisti (2005) also revealed that only 0.06% of 4,000 students at Carnegie Mellon university changed Facebook privacy settings to limit their profile visibility to strangers. The respondents did not know that Facebook gives them the ability to change the default privacy settings. In addition, 75% of 200 London Facebook users had their profiles viewable by other Facebook users who were not their Facebook friends (Sophos, 2007). Furthermore, only about 33% of 1,710 college students made their profile private in a study by Lewis, Kaufman, and Christakis (2008). Ninety nine percent of 67,000 Twitter users have also been reported to have retained Twitter’s default privacy settings whereupon a large volume of information on their profiles was accessible by the public (Krishnamurthy et al., 2008).

On the contrary, a later study by Young (2009) showed that the majority of Australian students tended to befriend only people that they knew offline and that they were conscious of the privacy settings for the visibility of their Facebook profiles. Similarly, the majority of Facebook users in
Debatin, Lovejoy, Horn, and Hughes (2009) report having an understanding of privacy settings and make use of those settings. Women in Saudi Arabia are also reported to take care with their SNS privacy settings as well as with the types of information they share with their SNS friends (Al-Saggaf, 2011). They tended to prevent those SNS friends (who were not their offline friends) from seeing their pictures because of cultural differences and their status as Muslim women.

There have been a number of reports which show that the majority of online users do not read the visited sites’ privacy policies. For example, Lawler and Molluzzo (2010) reported that 55% of 1,200 students at Pace University in New York did not read any SNS privacy policy, and they are not aware that their personal information was being gathered, used and shared to different parties or organisations. Similarly, in a study by Acquisti and Gross (2006), 77% of respondents did not read Facebook privacy policy and mistakenly believed that Facebook did not collect, or share their personal information with third parties.

6.4.2 Online tracking and user awareness

As discussed in Chapter 3, information about online users (identifiable and non-identifiable) that is diffused online can be tracked or collected, and combined in order to build a profile of interests that can be used in behavioural or targeted advertising (McDonald & Cranor, 2009). In order for users to understand the mechanism or the nature of online tracking or targeting advertisements, they need to have at least some basic knowledge or understanding of how HTTP cookies work. Hence, it is questionable to what extent online users are aware about the practice of online tracking.

The study by McDonald and Cranor (2009) reported that the majority of participants did not understand the use of HTTP cookies, nor did they know about online tracking, and they believed that their online activities were completely anonymous unless they were logged into a website. Participants were also reported to have very little understanding of how browsers or other technologies work in conjunction with HTTP cookies. A later study showed that participants found targeted advertising invasive and that some would change their online behaviour if advertisers were collecting data (McDonald & Cranor, 2010a).

However, participants in another study (Nguyen, Kobsa, & Hayes, n.d.) reported to show low level of concern regarding everyday tracking and recording technologies. However, it is not clear whether or not those participants were also aware of the secondary usage of their online activities. They may be unaware that their information can be used outside the context in which it is given (Staddon & Swerdlow, 2011). For example, some users may be aware that bits and
pieces of user information can be collected by third party sites, but they may not be aware that their information can be combined and processed by a variety of data mining techniques in order to build their online profiles or predictive models of their online behaviours that can be used to improve online targeted advertisements online.

The existing literature has suggested that users do provide a large amount of information about themselves to SNSs (Gross & Acquisti, 2005; Krishnamurthy & Wills, 2008; Young, 2009; Lawler & Molluzzo, 2010). At the same time, online users care and value their online privacy even though they did not read the privacy policies of the sites they visited (Acquisti & Gross, 2006; Lawler & Molluzzo, 2010). It is also observed that, over time, users became more aware of SNSs’ privacy settings and use them to limit access to their information by unintended audiences (other Internet users) (Young, 2009; Debatin et al., 2009; Al-Saggaf, 2011). The literature also suggested that although the majority of online users did not have a good understanding of online tracking mechanisms (e.g., HTTP cookies), they expressed concerns regarding the collection of their unintentionally shared information by numerous parties such as advertisers (McDonald & Cranor, 2009, 2010). However, the literature has not yet explored whether or not users are aware of the tracking or the collecting of users’ unintentionally shared information by SNSs.

Hence, this stage of this research explores this gap in the literature by not only gauging perceptions of online privacy, but also capturing participants’ reactions to the fact that their information can be tracked by major SNSs. The researcher conducted qualitative interviews aiming to determine participants’ online behaviour, particularly on SNSs (e.g., how much information they disclose), their perception and understanding of online privacy (e.g., online tracking in general) and their reaction to the tracking by SNSs based on the outcomes of their online browsing experiments. Chapter 7 details the methodology and techniques used to collect the interview data, whereas Chapter 8 presents the analysis and outcomes of the interviews with 20 Cambodian participants.
Chapter 7

Social network sites, privacy and research participants: study design

The outcomes of Chapter 4 and Chapter 5 revealed that the researcher’s online activities as well as the research participants’ online activities can be tracked and recorded by major SNSs like Facebook, Twitter, and Google Plus. As identified in Chapter 6, online users’ perceptions of tracking by major SNSs has not yet been covered in the literature. This stage of the research seeks to ascertain to what extent research participants are aware of the practice of online tracking and how they react to the fact that their online activities are being tracked by numerous third party sites, particularly major SNSs. This chapter provides details on the conceptual framework and methodology used in this stage of the research. Section 7.1 demonstrates the intent of this stage of the research while Section 7.2 describes the methodology and techniques employed in this study. Section 7.3 presents the procedures and steps in collecting and analysing the data. Meanwhile, the outcomes of this stage of the research are presented in Chapter 8.

7.1 Intent of research

As elucidated in Chapter 5, this research collected two types of data from Cambodian research participants: HTTP messages and qualitative interviews. While Chapter 5 provides details on the data collection, analysis and outcomes of participants’ online browsing experiments, the present chapter provides details on the methodology and technique used in collecting qualitative interview data from participants. The outcomes of the online browsing experiments are revealed to participants during the interviews. Once they are aware of the tracking, they are in a better position to express their informed opinions on the concerns regarding the practice.
Their reaction is also captured during the interviews. This stage of research aims to address the following research questions:

- What are SNS users’ views on online tracking and privacy?
  1. What does privacy mean from SNS users’ perspectives?
  2. To what extent are SNS users aware of online tracking in general, as well as online tracking by SNSs?
  3. What are SNS users’ reactions to the practice of sharing and tracking information online?

This current work is similar to the studies conducted by Acquisti and Gross (2006); Sophos (2007); Lipford et al. (n.d.), and Lewis et al. (2008) as it investigates users’ awareness regarding privacy settings and data collection online. It is also similar to the research conducted by Nguyen et al. (n.d.) for it seeks to understand users’ perceptions regarding information collection online. In addition, it is also closely related to the work by Wills and Zeljkovic (2011) in that it seeks to understand the participants’ attitudes regarding privacy in the context of online tracking.

However, the differences lie in the use of methods, techniques and context such as the Cambodian context of the present study. Previous works relied mainly on surveys or automatic tools to collect the data. In contrast, this stage of the current research employs a qualitative approach. It is influenced by the ethnographic method and collects data by interviewing research participants. This research does not intend to study the whole population of Cambodian online users, rather, it aims to delve into the perspectives of Cambodian online users who have been widely exposed to new technologies such as smart phones, computers and Internet services including SNSs. Section 7.2 briefly introduces the qualitative research paradigm as well as the tools and techniques used in this stage of this research. It also briefly describe the characteristics of ethnography.

### 7.2 Methodology

#### 7.2.1 Research paradigm

Two major traditions of research in the social sciences are known as positivist and interpretivist. Positivists believe that a single reality exists and that the world is deterministic with a collection of observable events and facts which are measurable (Trochim, 2006; Williamson, 2002). Science is seen as a way to get the truth to understand the world, and the purpose of science is to stick to what we can observe and measure (Trochim, 2006). Positivist researchers
apply the research methods used in the natural sciences to the social sciences (Williamson, 2002). This research tradition is based mainly on deductive reasoning which works from the more general to the more specific, beginning with the theory, a hypothesis to narrow down the topic, observations to address the hypothesis and ultimately leads to the confirmation (Trochim, 2006).

Positivism is linked with the quantitative research approach. In quantitative research, the investigator conducts a systematic empirical investigation of a phenomena in order to answer the research questions or to test the predetermined hypotheses (Ary, Jacobs, & Sorensen, 2010; Muijs, 2011; Creswell, 2012). Data are usually collected in the form of numbers or converted to numbers, and can be analysed using statistical, mathematical or computational techniques. The literature is a key to suggest the research questions and to justify the need for the research problem. Quantitative researchers use random sampling where representative individuals are selected and results generalised to a large population (Creswell, 2012).

Meanwhile, interpretivism accepts multiple realities which are socially and individually constructed in people’s minds (Williamson, 2002). People make sense of their own world individually (e.g., they develop their own meaning which is often different among individuals) and together (e.g., they socially construct reality). Constructivism is a key interpretivist paradigm (Williamson, 2002). Constructivists believe that each of us construct our view of the world based on our perception of it. Interpretivism is based on inductive reasoning which works in the opposite way to deductive reasoning, starting from a specific observation to a broader theory (Trochim, 2006).

Interpretivist tradition is primarily associated with the qualitative research approach which aims to gather an in-depth understanding of the social phenomena from the perspective of the participants (Ary et al., 2010). Qualitative research focuses on the beliefs, feelings and interpretation of the participants. Qualitative approach uses purposeful sampling whereby the researcher intentionally selects sites or individuals that best help to develop an in-depth understanding of the studied phenomenon (Creswell, 2012).

This stage of the research aims to provide an understanding of how Cambodian online users perceive online privacy and how they react to the fact that their online activities can be tracked or recorded. Participants’ perspectives on the meaning of privacy, their perception of online tracking, and their reaction to the tracking by SNSs can only be best interpreted in a descriptive way; hence, a qualitative approach will best address the research questions. Next, Section 7.2.2 examines the characteristics of ethnography.
7.2.2 Ethnographic method

Ethnography, the primary method of anthropology, is the study of ways of life of humankind and the cultural basis of their peoplehood (Patton, 2002). It may include rigorous research and combine different approaches, quantitative and qualitative, with emphasis on understanding a full range of social behaviour within a particular location, event, or setting (Pole & Morrison, 2003). This method aims to provide an in-depth understanding of people’s views, actions, and the settings where they are located, via the use of observation and interview techniques. Hence, ethnography is useful in cultural studies, women’s studies, literary theory, nursing, law, planning and industrial engineering (Tedlock, 2003).

Ethnographic inquiry takes it as a guiding assumption that a group of people who interact together for a period of time will develop a culture (Patton, 2002). The central focus of ethnography is therefore to provide description, analysis, and interpretation of a cultural or social group’s shared patterns of behaviours, beliefs, and language that develop over time (Pickard, 2007; Creswell, 2012). Within ethnographic research methods, researchers have no prior assumptions but seek to explore while allowing the theoretical framework to act as cognitive signposts (Pickard, 2007).

The term *field work* in ethnography refers to the researcher’s engagement and data collection in the field or setting in which the participants live or work, and where their patterns can be studied (Pickard, 2007; Creswell, 2012). To best understand the patterns of a cultural group, it requires that the researcher or the investigator have a prolonged engagement or be otherwise immersed in the context or culture being studied (Patton, 2002; Pickard, 2007; Creswell, 2012). In some cases, the investigator is or becomes part of the context or culture being studied.

Human interpersonal relationships usually create complexities and patterns that need to be studied and understood, and in order to best understand this, the researcher acts as the research instrument, and the human experience and situations are the main subject of ethnographic research (Pickard, 2007). The overall picture of an ethnography comprises of a collaboration between the research participants and the researcher as research instrument, and qualitative data collection techniques that can be used to process and interpret data to present the description of the context. Before engaging in the field or social group, the researcher should be prepared to ask questions while at the same time acknowledging that multiple realities exist among participants as well as within the researcher (Pickard, 2007).

Participant observation is usually the primary technique for data collection. However, a variety of other techniques such as interview, survey, as well as casual conversation, are also justifiable (Pickard, 2007; Creswell, 2012). Ethnographic data can be emic, etic, or negotiation (Creswell, 2012). Emic data involves information provided by the participants, whereas etic data involves
information about the participants’ perspectives interpreted by the ethnographer. Negotiation
data occurs when the information or data comes from an agreement between the researcher and
the participants in the study (Creswell, 2012).

This study is influenced by the method of ethnography and employs one of the ethnographic
techniques, the interview, in collecting data from participants. This method and technique guide
the researcher in investigating the research participants’ perspectives on privacy and online
tracking. They also help the researcher to describe, analyse, and interpret the shared patterns
of behaviours and thoughts among a culture-sharing group, in this case, participants’ online
behaviours and perceptions of privacy online. Section 7.2.3 further explores interviewing as
a technique, which was employed to elicit and gather data from participants after their online
browsing experiments.

7.2.3 Interview technique

An interview is a specific form of conversation or interaction, with a structure and purpose, be-
tween interviewer and interviewee from whom the knowledge is produced or extracted (Kvale,
1996; Liamputtong, 2009). This technique allows some degree of interaction between re-
searcher and participants in a study, and it also allows researchers to access in-depth infor-
mation that is specific to individuals (Pickard, 2007). The researcher interview differs slightly
from the conversation between equal partners because the researcher has control over the topic.
The type of interview depends on the nature of the research topic and the type of data needed
to answer the research question (Pickard, 2007). Other factors such as the researcher’s experi-
ence, the available time, and the number of participants, are also taken into consideration when
deciding on the type of interview. Interviews can be either structured or unstructured in nature
(Pickard, 2007; Patton, 2002).

Within the structured or standardised interviewing situation, the participants are asked the same
series of questions with a limited set or predetermined answers according to a prepared cod-
ing scheme (Pickard, 2007; Minichiello, Aroni, & Hays, 2008; Al-Saggaf, 2011). There are
two forms of structured interview: open-ended and fixed-response interview (Pickard, 2007).
In the open-ended interview, the participants are asked the same series of questions but they
are allowed to answer in the way they choose to share, whereas in the fixed-response inter-
view, the interviewees can only choose from a predetermined set of answers (Pickard, 2007).
Fixed-response interviews are seen to be similar to questionnaires. However, it still allows the
researcher to capture visual and oral clues from the participants; for example, whether or not
they look nervous (Pickard, 2007).
The unstructured or loosely structured interview, on the other hand, appears to be like a typical everyday conversation, and is reliant on the social interaction between the interviewer and the informant to elicit the information (Minichiello et al., 2008). This type of interview is used to gain a holistic understanding of the thoughts, feelings and points of view of the informants (Pickard, 2007). Open-ended questions are used to allow participants to tell their stories in their own words. Two approaches can be used to conduct unstructured interviews: the informal conversation and the general interview guide (Patton, 1987; Pickard, 2007). The informal conversation could be the most difficult form of data collection because it requires the researcher’s familiarity with the topic and high concentration in listening and responding to the participants in the immediate context, while at the same time steering them along the lines of research interest (Pickard, 2007). Meanwhile, in a guided interview, the researcher prepares a checklist to make sure that all the relevant points are covered while conducting the interview. This type of approach is less challenging than the informal conversation, but also allows the investigator to elicit information about specific topics (Pickard, 2007).

How an interview is conducted may vary. Interviews can be one-on-one when the researcher interviews and records answers from one interviewee at a time. This process can be the most time consuming and costly, but is ideal for participants who are comfortable enough in sharing their thoughts and opinions with the researcher (Creswell, 2012). The interview can also be conducted in a focus group where the researcher collects the data by interviewing a group of people. This works best when the interaction within the group provides the best information to the researcher (Creswell, 2012).

Interviews can also be in-depth. In-depth interviewing involves one-on-one and face-to-face interaction between the researcher and the participant that resembles talking among friends, except the researcher seeks to use the information from the interaction for the purpose of research (Johnson, 2002). This type of interview allows the researcher to explore deeper information than is found in surveys, or focus groups. This information usually involves personal matters such as experiences, decisions, and perspectives.

Thanks to technology, interviews can be conducted by telephone or online, via e-mail, or instant chat. This is usually done when it is not possible for the researcher to gather the participants for face-to-face interviews, or when the study time is short (Pickard, 2007). For example, when participants and researchers are in different countries, online video interviewing could be a better choice than face-to-face interviewing because it saves time and travel costs. When deciding on face-to-face, telephone or online interviewing, the researcher should choose the technique most suited to helping answer the research questions. Other considerations include ethical, political, or pragmatic factors that might affect the research outcomes and research participants. For example, when participants are not comfortable with face-to-face interviews,
the researcher may conduct interviews by telephone.

For this stage of the research, the open-ended structured interview was chosen as a data collection technique to enable each of participants to be asked the same series of questions (see Appendix D). Unlike questionnaires or close-ended questions where the answers are predetermined, open-ended questions allow the researcher to explore varieties of responses among the participants (Creswell, 2012). The researcher planned to do face-to-face and one-on-one interviews with the 20 participants; however, only one participant was ultimately interviewed face-to-face in a cafe. Nineteen phone interviews were subsequently conducted due to time constraints and the participants’ availability and convenience.

The researcher spent three weeks in Cambodia. The first week was spent organising the workshop venue, and contacting and confirming participants for the Wireshark workshop. Following the workshop, the researcher analysed 20 Wireshark files during the second week of her stay in Cambodia (see Chapter 5). After this, the majority of participants commenced work while others were away on holiday after their graduation. It was not possible to gather participants for interviews over the following weekend as the researcher had to return to Australia. In addition, in Cambodia, it is not considered appropriate for females to go out alone in the evening. Therefore, the researcher conducted telephone interviews with 19 participants after working hours (after 6pm) during the third week of her stay in Cambodia.

### 7.2.4 Sampling and recruiting

As mentioned in Section 5.3, only participants who met the research criteria were selected to participate in the study. Twenty participants were recruited from IFL in Phnom Penh city. All subjects initially participated in online browsing experiments (see Chapter 5), then were contacted by the researcher for the interviews. The researcher met all of participants face-to-face during the Wireshark workshop and experiments; therefore, she was able to establish rapport with study subjects.

### 7.3 Data collection

#### 7.3.1 Conducting the interview

The data collection process started with participants’ online browsing experiments, and the researcher used the results in interviews that aimed to gauge participants’ awareness and perception of privacy and online tracking. As noted in Section 5.6, the workshop took place at
Lotus Blanc conference venue. The workshop involved the 20 participants engaging in online browsing on their laptops while having their online activities recorded by Wireshark. The researcher collected and analysed the Wireshark files (see Chapter 5) and revealed the analysis to each of them during the interview.

As stated earlier, the researcher initially planned to conduct face-to-face interviews with 20 participants, but in the event only one participant was available for a face-to-face interview. The researcher made phone calls to allocate suitable times for the phone interviews with the other 19 research participants. Before each interview took place, the researcher ensured that the call quality was good and that there was no disruption such as noises for each party. The interviews were conducted in English. During the interviews, there were occasions when participants had difficulty in fully understanding the questions in English, or they had difficulty in answering in English, so the researcher used Khmer from time to time to help elicit the answers.

Investigators may choose to record interview data by using a tape recorder or a recording application installed on their smart phone or computer, or by taking notes during the interviews (Pickard, 2007). In this research study, the researcher chose note-taking over a recording application or device. Note-taking was convenient for this research study as every participant was asked the same series of questions, and the researcher had prepared forms (see Appendix D) onto which she could write participants’ answers. There could also be risks involved in voice recording the interviews. For example, if a recording device or application fails (e.g., if the batteries run down, the microphone is not set up properly, or the application freezes) the researcher will lose participants’ real time answers. In addition, some participants may feel uncomfortable sharing their opinions and may decline the use of a recording application or device. For these reasons, the researcher chose note-taking for the interviews in this research study. Before each interview started, the researcher informed participants that she would not voice-record the interviews, but she would take notes of their answers on paper. During the interviews, she listened attentively to participants and wrote down their answers in the relevant space within the list of questions (as shown in Appendix D).

The researcher conducted the interview with each participant by going through the same series of questions in a conversational way. She also used probes or subquestions that may not necessarily have been in the set of questions in order to allow participants to clarify their points or expand on ideas. At the end of the interview, some participants also asked questions regarding the researcher’s browsing experiences while some asked about her research because they were interested in the area of online privacy. At the completion of the interview, the researcher thanked each participant and invited them for a farewell lunch which took place on the weekend before the researcher’s flight back to Australia.
7.3.2 Analysis of the interview data

After the interviews, the researcher organised the notes, transferred them from handwritten to word processed files, and selected the application for the analysis. Notes taken from the 20 participants’ answers were typed into Microsoft Word files. The typed files were saved under each participant’s nickname, and were stored in a secure folder on the researcher’s computer. Again, only participants’ nicknames were used within this thesis.

After the data are organised and typed into files, researchers can choose whether to analyse the data by hand or use a computer (Creswell, 2012). The hand analysis of qualitative data requires that researchers read the data, make notes, and divide it into parts. A qualitative data analysis computer program does not analyse the data; rather, it facilitates the analysis by providing features for storing, organising and labelling or coding the data (Creswell, 2012). NVivo Version 9 was chosen as a tool to analyse the interview data (QSR, 2013). NVivo software supports qualitative and mixed methods research. It handles non-numerical data. It does not do the analysis, but allows the researcher to organise the contents of the interviews or survey responses before analysing them.

During the interview data analysis, the researcher followed the following procedures:

- Handwritten notes taken from each interview were transferred into Microsoft Word files and saved under each participant’s nickname.
- The 20 files were imported into NVivo.
- As the 20 interviews used the same questions in the same order, the researcher chose Auto Code in NVivo to organise the text. For example, all the answers for each question were grouped together in one place or one node. With everything in one place, patterns are easy to find.
- While reading through the answers to each question, the researcher started to code primary and secondary themes identified from those answers. Coding is the process of dividing texts into segments and labelling them to form descriptions and themes (Creswell, 2012).
- The researcher then discusses the findings based on the identified themes.

The researcher’s approach to analysing the data was through thematic analysis where she searched for themes and patterns within the data. Hence, while analysing interview data, the researcher identified the themes that help to address the research questions. Chapter 8 presents and discusses the outcomes or themes identified from interviews with 20 participants in Cambodia.
Chapter 8

Social network sites, privacy and research participants: the findings

Chapter 7 discusses the concept of qualitative research as well as ethnographic method, which influenced the approach to this research. It also details the tools and techniques employed to collect and analyse the qualitative data. Meanwhile, this chapter presents and discusses the findings from the interviews with the 20 participants after the online browsing experiments. Four main themes were extracted from participants’ answers: SNS usage, intentionally shared information, privacy, and unintentionally shared information. They are presented respectively in Section 8.1, Section 8.2, Section 8.3, and Section 8.4. Section 8.5 discusses these themes and the literature, whereas Section 8.6 summarises the findings of this stage of the research.

8.1 SNS usage

8.1.1 SNS(s) in use

Eleven of the 20 participants created their first SNS accounts in 2005, while the remainder did so in 2008. Hi5 was seen to be the first SNS used among the majority of participants. Other SNSs like MySpace, IFler, and Angkorone were also popular among participants prior to the existence of Facebook, Twitter and Google Plus. Every participant reported that they used SNSs, particularly Facebook, on a daily basis. Eleven of the 20 participants used Facebook only, while the remainder used Facebook and other SNSs like Google Plus, and Twitter.

Facebook was the most visited SNS among participants. For instance, Kosal said, “I visit Facebook all the time! I spent at least five hours a day on Facebook.” Meanwhile, Nary had her
Facebook account signed-in all day. She constantly checked her Facebook profile for updates regarding her study group as well as her work. Udom added, “I visit Facebook almost everyday, quite often during the week, which is about 60% of my time each week.” Besides Facebook, the participants, who have multiple SNS profiles, also browsed other SNSs (such as Google Plus, Twitter, and LinkedIn), but not as often as Facebook. For example, Tevy browsed her Facebook page everyday while she browsed her Instagram profile only once in a while. Similarly, Sitha mentioned that she rarely or almost never checks her LinkedIn profile because she does not have many connections (friends) there.

8.1.2 Motivation for creating SNS accounts

Seventeen participants began using their first SNSs because of influence from their peers. For example, Rany, Phary and Seyha signed up for their first SNS account because they received e-mail invitations from their friends who were already members of that network. Similarly, Moly said, “I signed up for a Hi5 account about three years ago because I got an e-mail invitation from my high school friend. Hmm. And I got to know Facebook from my classmates at university.”

Some participants were introduced to SNSs via different means. For example, Bopha first became aware of Twitter and signed up for a Twitter account after she saw Twitter buttons on many Web sites she visited, while Vibol got interested in Twitter and signed up for an account after seeing Twitter logos in various online ads. Four of the 20 participants also reported they signed up for an SNS account because they wanted to explore SNS features. For example, Dara, Kosal, and Rany signed up for a Facebook account because they liked the features that are available on Facebook (e.g., third party applications, and news feeds), but are not available on other SNSs. Meanwhile Mony signed up for a Google Plus account because he is interested in Google and its new products. Celebrities’ tweet messages also influenced a participant to join SNSs. For example, Nary said, “I signed up for a Twitter account because I’m interested in following the tweets of my favourite celebrities.”

8.1.3 Browsing activities and technologies

Besides using SNSs, participants also reported checking e-mails, performing online searches, and reading online news articles on a regular basis. They were involved with reading and writing articles on blogs. Meanwhile, Nara, used Internet for business purposes. He owned and operated a Web site that served as a Web portal. His site displayed local advertisements about
certain products (e.g., new and used mobile phones) from different local stores or individual sellers in Cambodia, but it does not involve any payment transaction.

All participants reported using their personal computers as a means of accessing the Internet. Eighteen also browsed online on their smart phones while 4/20 accessed the Internet via their iPads. Participants used different browser such as Google Chrome (12/20), Firefox (9/20), Safari (8/20), Opera (3/20), and Internet Explorer (1/20), depending on the device they used. Twelve participants reported using more than one browser. For example, Kosal browsed online using Safari on his iPad and Firefox on his laptop. Safari is the most popular browser among mobile phone and tablet users due to the fact that it preloaded on iOS devices (iPhones and iPads).

8.2 Intentionally shared information

8.2.1 Information shared on SNSs and motivation for sharing

Two categories of information are intentionally disclosed on SNSs by research participants: information required by SNSs during sign up, and other information or content they generate while using SNSs.

All participants shared information they perceived to be required by SNSs during the sign-up process. This type of information included e-mail address, full name, profile picture, date of birth, location, hobbies, educational background and work experience. They shared this type of information because they believed it was a requirement of the SNS and that SNSs would not allow them to create a profile if they did not provide this information. For example, Dalin said, “I was worried that I couldn’t sign up for a Facebook account if I didn’t provide as much information as required in the form during the sign up process.” Similarly, Rany mentioned that she was also worried that Facebook would not allow her to sign up for a profile if she did not provide enough information to Facebook. She also added that she does not think that Facebook would use her information for purposes other than helping her connect with her friends.

Meanwhile, participants also generated other content on SNSs while communicating with their SNS friends. This content can be general (such as games, external contents from other sites, music, pictures and videos) or personal (including feelings, personal achievements, and physical movements or locations).

Half the participants stated that they generated this content to stay in touch with their friends, and to allow others to get to know them better. For example, Mony said, “I share my educa-
tional background because I want to introduce myself to my Facebook friends.” Meanwhile, Phara shared a lot about himself so that his profile would be searchable by people he may know, especially by his friends. Similarly, Phanna and Rithy liked sharing information about themselves on SNSs in order to keep their friends up-to-date about them, which they believe is crucial for friendship.

The other half of them liked sharing their feelings on Facebook because they enjoyed receiving emotional support from their friends. Seyha said, “I do it because I like sharing. When I’m stressed, I share my personal feelings, I love to get comments from friends.” Similarly, Dara and Nary mentioned that they liked to express and share their feeling on Facebook because they liked getting positive and encouraging comments from their friends which eventually makes them feel better.

Some other participants reported sharing information on SNSs because they like to explore the SNS features and those features enhanced their experience in sharing information with others. For instance, Kosal mentioned that, “I love to explore different features of Facebook and blogs, especially the concept of sharing information with the world.” Similarly, Phary said that sharing pictures on Facebook is much easier than transferring pictures to a USB device and passing it around among her friends.

Seven out of 20 participants indicated that they liked to share what they are passionate about. For example, Tola shares the links to new songs or music videos because he has a passion for music. Likewise, Udom said, “I personally love reading philosophy from well-known people. Those people or philosophers love to share their words of wisdom, so I also wanna share the knowledge to my Facebook friends”. Similarly, Moly has a great passion for photography and fashion; hence, she shared many pictures relating to fashion and models.

### 8.2.2 SNS privacy settings

All participants in this study were aware of SNSs’ privacy settings. Sixteen of them had changed their privacy settings so that their Facebook profiles were private and accessible only by their friends. Public or non-SNS-friends could only see profile pictures and some very basic information such as gender and name. In addition, some of them also mentioned that they now no longer accept friend requests from strangers, but only from people they know in real life. For example, Sitha mentioned, “My Facebook profile is private. Only my friends can see what I post there. I no longer accept friend requests from people I don’t know, like I did before. I also removed contacts that I don’t know from my friend list, because I don’t know them. I don’t trust them.”
Of these 16 participants, 7 have their other SNS profiles (e.g., Google Plus and Twitter) accessible to the public. They chose to make their profiles public or did not care about changing privacy settings claiming they did not share much information about themselves on those SNSs. For instance, Bopha set her Facebook profile to private, available to her friends only, while her Google Plus and Twitter profiles are available to the public, because she claimed she never posts anything about herself on Google Plus and Twitter.

Four out of 20 participants had all of their SNS profiles available to the public, claiming they did not disclose many personal details. For example, Udom mentioned that, “I used to be very careful about who can view my profile, but now I don’t care, I can’t be bothered to restrict too much. I made my profile public, but I removed my personal details and I share less about myself. If there’s anything important, or personal, I PM [private message] my friends.”

### 8.2.3 Privacy policies

In relation to SNSs’ privacy policies, 17/20 participants stated that they never read privacy policies at all because they find these policies to be too long, boring and too complicated for them to understand. Vibol said, “I’ve never read any privacy policy because it is too complicated. I just clicked Yes-Yes-Yes, in order just to proceed and to get an account.” On the other hand, two participants reported reading SNSs’ privacy policies. They commented that those policies are too long, not clear enough and very confusing for them to understand. Udom said, “Privacy policy does not look complete, not very clear. They should make it clearer and more simple for normal users. Yeah, it should be very clear.” Likewise, Seyha said, “I used to read the policies, just a little bit, not much, and I’m not sure what they are for.”

The majority of participants mentioned that they care more about SNSs’ privacy settings than what is stated in the privacy policies. Some of them also commented that the privacy policies gave them no choice but to agree in order to proceed. For instance, Rany said, “I’ve never read the privacy policy, because it is too long [laughing]. I’ve never thought about any problem related with the security. Even though I did or did not read the policy, I still have to click Yes in order to create an account.”
8.3 Privacy

8.3.1 Definition of privacy

Privacy was perceived by some participants as one’s control over the distribution or sharing of one’s information. For example, according to Makara, privacy is, “Anything about us that we have control on. It is something no one could take away from us without our permission.” Similarly, Nary defined privacy as: “Information that we have control on. We can decide whom to share that information with.”

Privacy was also perceived by participants as one’s right to one’s information whether or not that individual decides to disclose it to a specific group of people. Phara said that, “Privacy refers to something related to personal stuffs of a specific person. It should be respected. That person has a right to whether or not share it to others.” Meanwhile, participants also perceive privacy in a sense of others’ limited access to an individual’s information. For instance, Phanna defined privacy as “a state where not many people know about me.” Similarly, Rany said that, “To me, I think privacy is about having no interference or access from others on one’s private information.”

The meaning of privacy from the perspective of Cambodian participants revolves around the notion of an individual’s right, and/or control on others’ limited access to his/her information.

8.3.2 Personal or private information

What was considered private or personal varied between individuals. Family matters or details about one’s family, academic performance and relationship matters were considered personal or private for almost every participant. For example, Moly said, “I consider family matters and my academic performance as my personal matter. I never talked about my family issues on my Facebook wall.”

Medical records, financial situation, physical movements, and one’s business or future plans were also considered private among research participants. For instance, Kosal mentioned, “What I consider personal or private and what I don’t want to share to the public is my relationship and family matters, my medical records, my banking details, and my business plans.” Similarly, Mony mentioned that, “Almost everything should be private to me, because I don’t like to share much. Hmmm, physical movement should also be private, as well as my search keywords on Internet.”
Having said that, the majority of the participants indicated that the diversity in relationships or friendships defined what is private, because there were certain things that they would only share with certain groups of people such as friends, acquaintances and family. When asked, Tola and Rithy said that it was hard for them to explicitly define what is personal or private because they like to share different things with different people depending on their relationship. Although Phara is very cautious about his physical movements, he is expected to inform his family where he is and what he is doing. In addition, while participants do not talk about private matters on their SNS profiles, they discussed these with their friends on SNSs’ private messages.

### 8.3.3 Importance of privacy

Participants showed the value and importance of their privacy by expressing concerns about privacy loss and its consequences. To them, privacy loss could lead to embarrassment, discrimination, unfairness, and worse, safety issues. They also raised concerns that these consequences will also eventually have negative impacts on future opportunities as well as emotions and feelings. Phara mentioned that there was a time when his friend posted his exam results on his Facebook wall and that as his mark was not very good, he was very embarrassed and unhappy that his Facebook friends saw that post. Udom added that, “Privacy loss will become one’s weakness. It can haunt that person in the future.” Meanwhile, Phanna would feel very uncomfortable if a stranger knew his physical movement.

Privacy loss also influences decision making. For instance, Sitha mentioned that, “Privacy is quite important! If they know too much about me, they will probably gossip, or even interfere with my decision making.” Makara responded that privacy appears to be overlooked in his family in daily life. This is due to the fact that he has grown up in a family where he shares his sleeping and study space with his siblings and where he has to inform his parents about his daily activities (e.g., where he would go or who he would be with). However, he expressed that he considers and keeps his online activities and his diary very private.

### 8.4 Unintentionally shared information

#### 8.4.1 Views on privacy and online activities

Fourteen of the 20 participants believed that their online activities remained absolutely private and inaccessible by others so long as they did not disclose this information to others or that they browsed online in a private location, such as their room. For example, Dara said, “Well,
everything I do online is private because I never told anyone else about what I search on Google, or read online.” Meanwhile, 4/20 participants were not sure if the Internet was a private space. For instance, Moly said, “I’m not sure if what I do online is private or not. I am not a tech-savvy user. Maybe our online activities are not private because I’ve heard on the news about hacking, or stalking online. I also heard that the United States government can know what the criminals do online.” Two participants believed that Internet is a public space and that there is no privacy in what we do online. Nara said, “I don’t believe in privacy online. Everything we do online is not private.” He told the researcher that while he is a lecturer at IFL, he also runs a Web portal site and has a good understanding of how HTTP cookies work.

Of the 20 participants, 6 added that while they think their online activities may or may not be private, their activities on SNSs are not private at all, because other friends on SNSs may see them. For example, Phary said, “Some online activities are private but some are not. I believe Facebook messages are private, but what I post on my profile is not private because my friends can see it. I also believe that my online search via Google is private because there is no name related with the searches.”

8.4.2 Views on tracking by third party sites

The researcher revealed some elements of the outcomes of the online browsing experiments to participants. The outcomes indicated that participants’ online activities were being monitored and tracked by third party sites like advertisers and data aggregators for different purposes such as targeted advertisements, as well as other unknown purposes. The researcher explained about first and third party sites, and targeted advertisements, and found that 19/20 research participants were not familiar with these terms and were not aware of any type of online tracking, let alone the tracking by third party sites. Rather, they perceived online tracking to be the tracking done by other Internet users who may have had access to their information posted on SNSs.

Sixteen participants indicated that they were surprised and upset that their online activities were not private and were known by third party sites of which they had never heard. They also expressed concerns about privacy and deprivation of freedom. Seyha and Vibol stated that they had lost their freedom and privacy in online browsing when they learnt their activities could be tracked by advertisers like Google, just because they were using many services provided by Google such as e-mail, SNS, blogging, and the search engine. Both Seyha and Nary added that they will be more cautious regarding what they do online.

Some participants responded that they were fine with online tracking for a number of reasons.
Tola mentioned that, “I think as long as what we are doing online is clean, there shouldn’t be any problem of others tracking or knowing our online activities.” Phanna also added that, “I am okay with this tracking because even though my online activities are being tracked, I don’t think anyone really cares about me. I am not a famous person [laughing].” In addition, Dara and Kosal are fine with the tracking because they enjoy the relevant advertisements on their visited pages. Dara said, “Online tracking is not a big deal. I don’t see any problem related with relevant advertisements.”

8.4.3 Reaction towards tracking by SNSs

At this stage, the researcher also revealed that participants’ online activities can be tracked not only by various unknown advertisers or data aggregators, but also by major SNSs like Facebook, Twitter, and Google Plus. The researcher also added that SNSs acquire not only information provided or generated by participants while signing up for SNS accounts or while interacting with other SNSs users (their intentionally shared information), but also information about their online browsing among sites embedded with SNS widgets such as Facebook’s Like button, Twitter’s Tweet button or Google’s +1 button (their unintentionally shared information). This means SNSs can have a complete profile of an individual’s life.

All of the participants were surprised after learning that SNSs knew so much about them and their online activities. Eighteen participants stated that they were unhappy about such a practice. Some also expressed that they were concerned about others’ judgements about their personalities based on their online activities, and some became paranoid about what they do online. For example, Vibol and Mony indicated that they felt uncomfortable about their online activities knowing these can be tracked by giant companies like Google and Facebook. Vibol was concerned that he did not know what will happen to his profile or information. Likewise, Setha felt as if she was being stalked or followed by paparazzi. Phary added that one’s online reputation might influence others’ judgement of that individual, and what is worse is that a person’s online reputation does not accurately represent his or her personality. Similarly, Dara and Udom commented that this practice is not fair and could also affect one’s future opportunities if combined information could be sold to insurance companies or employment recruiters.

The majority of participants also expressed concerns about the privacy loss and privacy violation of this practice, and stated that users’ permission is crucial in this situation. Tevy felt that she had lost her privacy since her online activities were known by SNSs. In addition, Phara and Mony commented that the act of collecting information by SNSs without users’ informed consent is not right and clearly violates users’ privacy. Furthermore, Seyha and Tola added that the tracking companies should at least seek users’ permission before they track users’ online
activities. Phanna said that, “Oh really? I never knew this is happening. I feel sad - that hurts me - how can Facebook collect so much information about me without my permission?” Similarly, Seila added, “It feels like I am sharing my personal stuff to the whole world - there’s no privacy. Everything is for sharing - I am angry and unhappy about this. Maybe it’s wise to be more careful with what I’ll be doing online.”

In contrast, 2 of 20 participants were fine with the current practice, claiming that good profiles would yield good online reputations and that the right advertisements could also be beneficial for individuals. For instance, Kosal said, “I am fine with online tracking. I think it is convenient that the companies are able to provide us relevant ads. I am cool with online tracking.” Similarly, Nara mentioned that he is fine with this tracking but he is more concerned with the security of his information such as his password and the contents of his private messages on SNSs.

8.4.4 Protection from online tracking

With regard to protection from online tracking, 19/20 participants were not aware of any tools or technologies that could be used to lessen the chance of tracking by third party sites. As discovered in Section 8.4.2, they were not aware of tracking by third party sites until they were interviewed by the researcher.

After the interviews, some participants were willing to take some action to reduce the chance of tracking. For example, Seila stated that he will sign out of his SNS profiles every time he wants to visit other Web sites in order to prevent the tracking by SNSs. In addition, Rany mentioned that she will regularly clear cookies and browsing histories in her browsers because it helps to reduce the chance of tracking and linking to some degree, and that she learned about this from the Wireshark workshop. However, she commented that it would almost be impractical to take this approach when she has many websites to visit in one browsing session. For the time being, Kosal mentioned that he would be more careful with his online activities as well as with what he shared online, knowing that the Internet is not a private space.

Meanwhile, Bopha indicated that she wanted to know what type of protection was available on other devices like smart phones and tablet devices. Furthermore, Udom suggested that there should be a solution to the issue, either in the form of a warning or alert to the users or in the form of education about the current practice of online tracking, particularly tracking by SNSs. Phanna is the only participant who was using Firefox’s ad blocker tool in order to reduce the number of advertisements on the page. However, he was not very sure how effective that tool could be in reducing tracking.
8.5 Discussion: the findings and the literature

The findings of this study appear to be consistent with the literature to a certain extent. For example, Facebook is the most popular SNS among participants as well as among other users. Communication with peers is the key factor that encouraged Cambodian participants to start using SNS services, similar to the participants in previous works (Acquisti & Gross, 2006; Lewis et al., 2008; Joinson, 2008; Pempek et al., 2009; Young, 2009; Ross et al., 2009; Al-Saggaf, 2011; Cheung et al., 2011). Also comparable to other participants in previous works (Gross & Acquisti, 2005; Young, 2009; Debatin et al., 2009; Lawler & Molluzzo, 2010; Taraszow et al., n.d.; Brandimarte, Acquisti, & Loewenstein, 2010; Al-Saggaf, 2011; Staddon, Acquisti, & LeFevre, 2013; Hazari & Brown, 2013), participants in this study disclosed a large volume of information about themselves on SNSs, and all of them were aware of SNS privacy settings. Their intentionally shared information can be factual (e.g., full name, e-mail address, and date of birth) and information that is related to their interests or hobbies (e.g., favourite books or music) (Gross & Acquisti, 2005; Taraszow et al., n.d.; Lawler & Molluzzo, 2010). Participants’ sharing of information is also necessary for them to establish their online identities, stay in touch and interact with their SNS friends (Livingstone, 2008; Pempek et al., 2009; Waters & Ackerman, 2011).

Previous works (Gross & Acquisti, 2005; Acquisti & Gross, 2006; Taraszow et al., n.d.; Tad-dicken, 2014) have also showed how much information users shared on SNSs, but overlooked the key difference in regard to the users’ intention to share - whether or not they voluntarily do so or because they think they are required to. In order to sign-up for a new account, a new user is required to provide a minimal amount of information including first name, last name, e-mail address, and date of birth. Other information such as location, profile picture, educational background, work experience and interests are just additional content that is not compulsory to disclose, but necessary to help others get to know a user better. Some users, like participants within this study, share those details because they believe that that extra content is also required to initially sign up for an account.

In accordance with the existing literature (Barocas & Nissenbaum, 2009; Mayer & Mitchell, 2012), participants in this study found privacy policies too long and complicated for them to understand and the majority of them did not read these policies. In addition, they also raised the issue that privacy policies give them no choice but to agree in order to sign up for an SNS account. Language could also be a barrier because none of the policies is written in Khmer. Participants’ perspectives on the meaning of privacy appears to be consistent with existing privacy theories that perceive privacy to be an individual’s right (Warren & Brandeis, 1890),

1http://www.alexa.com/topsites
and/or control (Fried, 1990) over the limited access (Gavison, 1983) by others to his/her information. Cambodian participants treated their financial situation, academic performance, family and relationship as personal or private matters. The importance of privacy was expressed via the concerns about privacy loss which could lead to embarrassment, discrimination, unfairness and safety issues (Fried, 1990; Kang, 1998).

In addition, existing literature has not discussed the relationship between the motivation for sharing information online or the perceptions of privacy and culture, particularly Khmer culture. As briefly mentioned in Chapter 2, Khmer children are advised or taught by the didactic codes (or Cha-bab) not to talk about family matters to others, as it may bring embarrassment to the family. Khmer culture is also very hierarchical. Khmer children are taught to be polite, obedient and pay respect to parents and elders. In addition, the choice of spouse in Cambodia is a different process, compared with western countries. Although some parents in larger cities are accustomed to their children choosing their own spouses, parents from rural areas still prefer the practice of arranged marriage. Dating and premarital sex is a taboo (FRD, 1987). The family’s honour, reputation or face is highly valued in the society and is viewed to be in the hands of offspring of the family. For example, if children perform well at school or university, and have a professional career, they are seen to bring pride to the family. In contrast, they would bring shame to the family if they are disobedient to their parents. All participants disclosed information about their educational and working backgrounds on their SNS profiles. In contrast, although relationship or friendship defines what is private or personal, participants can be sensitive to the topic of family matters, relationships, academic performance and financial situation, and would not be comfortable in discussing these matters on SNS profiles that could be seen by others.

In regard to online tracking by third party sites, the majority of participants in this study were not aware of the practice. This finding is in line with the study by McDonald and Cranor (2009). However, 71% of the participants in the study by TRUSTe (2008) claimed to be aware of online tracking by third party sites for advertising purposes. Participants in the present research perceived online tracking to be the tracking done by other Internet users rather than by technologies like HTTP cookies. They also believed that their online activities were absolutely private as long as they did not reveal their activities to others.

Once they learned that their online activities were being recorded by third party sites, 16/20 participants became very surprised and expressed concerns about the loss of privacy and freedom in browsing online. This finding is also mirrored in studies by Turow, King, Hoofnagle, Beakley, and Hennessy (2009); McDonald and Cranor (2010a); Morales (2010); Wills and Zeljkovic (2011) and Purcell (2012) where the majority of participants show concern about third party sites monitoring online activities and do not wish to be tracked. Later, when the
researcher revealed that major SNSs like Facebook, Twitter and Google Plus are also tracking their online activities, 18 participants expressed concern about privacy loss and others’ judgments of their personalities based on their browsing activities. They were very unhappy that SNSs track their activities without asking their permission, even though this practice was stated in SNSs’ privacy policies which were overlooked by participants as well as other online users. Two of the study subjects were fine with the practice, claiming good profiles would be like good CVs. The study of tracking by SNSs and online users’ reactions or perceptions is relatively new and has not previously been discussed in the literature.

Looking specifically at the Cambodian context, extended families are very common in that nation. Family members usually share sleeping and study space. Cambodians are also taught to be open and transparent to their parents. For instance, they have to inform or ask permission from their parents before going out. Thus, it can be difficult to keep certain things private from family members. Privacy in Khmer society is commonly perceived as personal matters. There is a constitutional right to privacy which refers to “The right to privacy of residences, and to the secrecy of correspondence by mail, telegram, fax, telex and telephone shall be guaranteed” (Anonymous, 2012). However, there is no online data protection or privacy law in the country. Issues on privacy in general or online are not commonly discussed in the Cambodian media either; hence, the majority of participants were not familiar with the tracking by third party sites. Despite all of these factors, participants in this study reported that they valued their privacy. They found tracking by SNSs (without users’ permission) intrusive to their privacy and freedom in browsing online. As determined in this study, participants disclosed a large volume of personal information on their SNS profiles. However, this sharing is for their intended audience such as friends and relatives, but not for SNSs to combine and link with their browsing activities.

8.6 Summary

This stage of the research gauged Cambodian participants’ perceptions of privacy and online tracking after their online browsing experiments. This is a qualitative piece of work and is influenced by ethnographic method. The sample size is relatively smaller than that of a quantitative study. In addition, this study does not aim to generalise to the whole population in Cambodia. However, it sought to understand why participants have the views they hold, and their motivation behind the actions they take online, including the sharing of information on SNSs. Open-ended structured interview was chosen as a data collection technique. This technique gave participants freedom in expressing their opinions. It also allowed the researcher to engage in the conversation while also eliciting the answers.
Participants were recruited from IFL in Cambodia. They first took part in a Wireshark workshop, and then were part of an experiment in which they browsed online while having their activities captured by Wireshark. The analysis of these online activities shows that participants’ online activities and their information can be tracked by third party sites which are advertisers or data aggregators, and SNSs. These results are necessary for the researcher to assess participants’ awareness regarding online tracking, and to capture their reactions and feelings during the interview.

Meanwhile, the auto-code feature in NVivo allowed the researcher to organise the interview data based on questions and extract the themes from each of the answers of the interview questions. There are similarities and differences in these research findings compared to the existing literature. For example, mobile phones and computers are popular devices with which to access the Internet, and the most popular online activity among the participants is using SNSs (Brenner, 2013; Brenner & Smith, 2013).

The majority of participants reported using SNSs since 2005 when they commenced their university degrees. Even though they used more than one SNS, Facebook is by far the most used SNS and all participants used Facebook on a daily basis. Communication among their peer group was seen to be the main motivation for them to sign up for an SNS account, while some of them first decided to join SNS because they received e-mail invitations from their peers, without necessarily having any prior intention to communicate.

Information disclosed by the participants on SNSs, or their intentionally shared information, includes factual information (e.g., name, e-mail address, pictures, and date of birth), information about their interests (e.g., favourite movies, and books), and other information generated while using SNSs (e.g., comments, and feelings). The majority of participants shared this type of information because they thought they were required to do so in order to sign up for an SNS account, while others enjoyed the sharing and interaction or sought emotional support from their friends.

Privacy from the perspective of Cambodian participants revolves around the notion of individual’s rights, and/or control in regard to others’ limited access to their information. Participants also held different views regarding what is considered to be personal or private information. The majority of participants considered family matters, academic performance, and financial details as private matters. Having said so, participants noted that relationships defined what is private, because there is some information they can disclose to their close friends but not to the public. Personal information from participants’ points of view is seen to be associated with Cambodian culture in the sense that a family’s reputation is considered to be of social value, and Khmer children are taught from a young age that family matters should not be disclosed to
others outside the family. Children’ poor performance at school or low income could also lead to embarrassment in Cambodian society. Participants also expressed fear or concern about loss of privacy which leads to embarrassment and misjudgements about them by others.

The majority of participants did not read privacy policies online, claiming they had no choice but to agree with all the terms and conditions. They also mentioned that those policies were usually very long and difficult to understand. To the researcher’s knowledge, there is no privacy policy written in Khmer which is Cambodia’s official language. Regarding the usage of SNSs, participants were aware of SNS privacy settings. The majority intentionally changed privacy settings in order to limit their audiences while some did not restrict their settings because they claimed they did not share much information about themselves on SNSs. The majority of participants perceived that what they posted on SNSs was not private due to the fact that other SNS users (e.g., their friends) could view these posts. However, they considered browsing online or performing online searches were private as long as they did not tell others what they did. The majority of subjects were not aware of tracking by third party sites, or by SNSs.

All of the participants were very surprised when the researcher revealed the results of their online activities which involved leakage of information from first party sites to third party sites which are advertisers, data aggregators, and SNSs. Eighteen participants expressed their concerns and sadness about the current practice of online tracking, saying those tracking companies (e.g., Google, Twitter and Facebook) should have asked their permission. They also expressed concern about privacy loss and misjudgements about them by others based on their online activities. However, a small number of them were not displeased with the practice and claimed that good digital profiles yielded good online reputation, and that the right advertisements could be beneficial for users. Nineteen participants were not aware of any tools or technologies available to reduce the likelihood of online tracking. Only one participant reported using Firefox ad blocker to reduce the advertisement content on the sites he visits.

In the next stage of the research, the analysis is made of the practice of tracking users’ unintentionally shared information by SNSs, by employing a CI framework in order to understand privacy expectations and implications from an ethical viewpoint.
Chapter 9

Users’ unintentionally shared information: a moral perspective

This stage of the research analyses from a moral perspective the practice of tracking users’ online activities or unintentionally shared information by SNSs. The analysis employs a Contextual Integrity (CI) conceptual framework in order to evaluate the flow of information between SNS users and SNSs. This chapter also examines other available frameworks such as Restricted Access/Limited Control (RALC) theory (introduced and expanded by James Moor and Herman T. Tavani), and the ontological interpretation of information privacy (introduced by Luciano Floridi). However, only CI (introduced by Helen Nissenbaum) is chosen for the analysis in this thesis. The chapter is organised as follows. Section 9.1 highlights the key findings from earlier stages of research while it also justifies the need for the analysis from a moral perspective. Privacy theories such as RALC, and the ontological interpretation of informational privacy as well as CI are examined in Section 9.2. The CI framework is applied to the analysis in Section 9.3. Section 9.4 concludes the analysis as well as this stage of the research.

9.1 Summary of the findings

The investigation of HTTP headers in the exploratory study as well as with the 20 Cambodian participants indicated that online users’ unintentionally shared information is available to and can be collected by numerous third party sites. The third party sites in this research were identified to be advertisers or data aggregators, as well as SNSs. While other outcomes from the experiments are equally important, this research focuses only on the tracking of users’ unintentionally shared information by major SNSs.
Meanwhile, the findings of the interviews revealed that Cambodian participants did care about privacy online and in general. All subjects were aware of privacy settings among SNSs and a majority changed the privacy settings to limit their audience. They were cautious about what they posted or shared online, particularly among SNSs. The majority had never read privacy policies of the visited sites while 2/20 participants put effort into reading the privacy policies but expressed that they had difficulty in understanding the contents. In their opinion, their online browsing was private in the sense that no other person knew what they browsed online. However, they were not aware of online tracking by major companies like advertisers or data aggregators, let alone the possible tracking by SNSs like Facebook, Twitter and Google Plus. In this case, in addition to holding a large volume of users’ intentionally shared information, Facebook, Twitter and Google Plus are now also collecting and storing users’ unintentionally shared information. Users, particularly the participants within this study, were not aware that their browsing activities could be tracked by SNSs.

Based on these outcomes, the researcher is interested in determining whether the practice of tracking users’ unintentionally shared information by SNSs violates users’ right to privacy from a moral viewpoint. Technical accounts of privacy usually do not explain why privacy matters, or what is wrong with breaching it on moral grounds. In addition, the practice of tracking by SNSs is relatively new and has not yet been discussed from a moral viewpoint in the existing literature. This stage of research capitalises on the strengths of both approaches (technical and moral). The researcher believes that analysing an issue from a philosophical or moral perspective should be supported by the results of empirical research. The next section examines three available privacy theories.

### 9.2 Benchmark theories of informational privacy

This section briefly examines benchmark theories that have recently been put forward to address controversies affecting informational privacy. Three theories are outlined here: (1) RALC/ or Restricted Access/ Limited Control theory, (2) an ontological interpretation of informational privacy, and (3) Contextual Integrity.

#### 9.2.1 RALC framework

Moor (1990, 1997a) and Tavani (2007b) introduced and expanded upon a model of privacy known as Restricted Access/ Limited Control or RALC theory of privacy. An account of the concept of privacy, an account of the justification of privacy and an account of the management
of privacy are the three main components of this framework (Tavani, 2007b).

According to RALC, “An individual or group has normative privacy in a situation with regard to others if and only if in that situation the individual or group is normatively protected from intrusion, interference, and information access by others” (Moor, 1997a, p. 30). The notion of situation or zone is used to determine or justify whether information should be normatively protected, and the notion of privacy is defined in terms of protection from intrusion and information access by others in the context of a situation (Tavani, 2007b). The theory also distinguishes between the condition of privacy and a right to privacy. This distinction should help to differentiate between a loss and a violation of privacy (Moor, 1997a; Tavani, 2007b).

Unlike the case of a normatively private situation where there are established laws and norms to protect one’s right to privacy, there are no conventional, legal or ethical norms according to which one has the right to be protected in the naturally private situation. Hence, privacy can be lost in a naturally private situation whereas in a normatively private situation, one’s privacy could be violated or invaded, in addition to being lost. “An individual X, may indeed lose some privacy (in a descriptive sense) whenever data about X is accessed” (Tavani, 2007b, p. 14). For example, Jane’s privacy can be lost to some extent when others happen to see her working on her assignment in a computer lab at her university. However, her privacy can be violated if someone peeps through her bedroom window and sees her working on her assignment.

In addition, in managing one’s privacy, an individual needs to have some degree of control with respect to choice, consent, and correction. “A person needs some control in choosing situations that offer others the level of access the person desires, which can range from total privacy to total publicity” (Tavani, 2007b, p. 12). One can also consent to others accessing certain kinds of information about oneself. On top of that, an individual should be able to access his or her own information and to amend it if necessary. These limited controls (choice, consent, and correction) can be enabled by adequate privacy policies (Tavani, 2007b).

### 9.2.2 An ontological interpretation of informational privacy

In Floridi’s framework of Information Ethics (IE), he argues that information has its moral worth and therefore, deserves moral consideration (Floridi, 1999). On the basis of biosphere, which refers to a limited region on the planet that supports life, Floridi coined a term known as the infosphere which refers to the whole informational environment constituted by all the information entities and their properties, processes, interactions, and mutual relations (Floridi, 1999, 2007). This environment or sphere includes online, offline and analog spaces of information; hence, cyberspace is only one of its subregions (Floridi, 2007).
According to his theory, “informational privacy is a function of the ontological friction in the infosphere” (Floridi, 2005, p. 187). This ontological friction refers to “the forces that oppose the information flow within (a region of) the infosphere” (Floridi, 2005, p. 186). For example, two people are located in different rooms and are separated by the walls. In this case, the walls are seen to provide high ontological friction for information flow between them. Floridi also draws a distinction between old or pre-digital ICTs (Information Communication Technologies) and the new or digital ICTs. In his views, ICTs are reontologizing the infosphere. Reontologizing refers to “a very radical form of reengineering, one that not only designs, constructs, or structures a system (e.g., a company, a machine, or some artifact) anew, but that fundamentally transform its intrinsic nature” (Floridi, 2007, p. 60).

He argues that the old ICTs have the ability to reduce the degree of informational privacy within the infosphere because they reduce the ontological friction. Within the same example of two people in different rooms, the obstacles in the communication or information flow between both of them can be enhanced or reduced by using a telephone. On the other hand, the new ICTs have the ability to not only decrease, or protect informational privacy, but also to alter the nature of the information (Floridi, 2005). For example, large volumes of information can be collected, processed, and distributed rapidly in the digital age, regardless of physical distance. Technologies like encryption also enhances the security of information storage and access.

A person S’ informational privacy refers to “S’ freedom from epistemic interference or intrusion, achieved thanks to a restriction on facts about S that are unknown or unknowable” (Floridi, 1999, p. 52). “Privacy is nothing less than the defence of the personal integrity of a packet of information, the individual; and the invasion of an individual’s informational privacy” (Floridi, 1999, p. 53). Floridi argues that a person is constituted by his or her information, and this nature allows one to understand the right to informational privacy (Floridi, 2005). His ontological interpretation suggests that “You are your information”, and that one’s infosphere is one’s identity. A violation of informational privacy is seen to be a violation of the nature of the information, but not a violation of ownership because informational privacy is not one’s belonging and personal information is a constitutive part of a me-hood (Floridi, 1999, 2005). The information flow also requires some friction in order to preserve the distinction between the multi agent system or the society, and the identity of the agents or the individuals that constitute the society (Floridi, 2006).

Floridi also formulated a qualitative equation that can be used to analyse the relation between the digital ICTs and informational privacy: “Given a certain amount of personal information available in (a region of) the infosphere I, the lower the ontological friction in I, the higher the accessibility of personal information about the agents embedded in I, the smaller the informa-
tional gap among them, and the lower the level of informational privacy implementable about each of them” (Floridi, 2005, p. 187).

9.2.3 Contextual Integrity

Nissenbaum (1997, 1998) argues that in order to address the challenge in the information technology age, theories of privacy should take into account the information which can be harvested in a public realm. She believes that Privacy in public, which has been ignored in the previous privacy theories, is worthy of being studied and protected in the information age. Nissenbaum is also opposed to two misleading assumptions: (1) “There is a realm of public information about persons to which no privacy norms apply” and (2) “An aggregation of information does not violate privacy if its parts, taken individually, do not” (Nissenbaum, 1997).

Regarding the first assumption, she argues that public space (e.g., a public park or sidewalk) is governed by some norms of privacy and that information collected from those spaces is not completely public. For example, when asked for their name by a complete stranger in the public park, an individual is entitled to answer “None of your business”. Moreover, she also points out that an individual has the right to control the dissemination of personal information even if it happened in a public area. For instance, even though a rape case occurred in a public park, the rape victim as well as her family retain some measure of control over the information about them - particularly her identity.

In response to the second assumption, Nissenbaum indicates that collecting, compiling, and combining innocuous pieces of information about an individual can be invasive of their privacy because those pieces of information can reveal a rich portrait of that person. Technology has enabled us to collect large amounts of information in a short period of time and to draw meaningful inferences from them which can be embarrassing to an individual. She also adds that the act of compiling information which involves moving information across different contexts results in information being used in a manner that is not explicitly known when it was initially collected. Therefore, this act failed, as permission to move their information around was not obtained from users.

Given that previous philosophical and legal theories of privacy offered limited justification for dealing with the problem of privacy in public, Nissenbaum introduced Contextual Integrity (CI) as a privacy benchmark (Nissenbaum, 2004). CI focuses on the notion of a context to analyse whether or not a flow of information, either gathering or dissemination, is appropriate with the norms within a context. Contexts refer to “structured social settings with characteristics that have evolved over time (sometimes long periods of time) and are subject to a host of causes
and contingencies of purpose, place, culture, historical accident, and more” (Nissenbaum, 2010).

Almost everything we do including acting and interacting with others takes place in contexts including education, health care, employment, family, business, and religion. People also navigate across different contexts throughout the day (Nissenbaum, 2010). For example, Jane is a PhD candidate at Charles Sturt University. She is also working as a part time web developer for a private organisation. She spends time watching television with her family, goes to her office on a daily basis, goes to work every Tuesday and Thursday afternoon, goes shopping, seeks medical care every six months, and goes to concerts with her friends.

Each context is also characterised by different roles, is partly constituted by canonical activities that are oriented around values, and is governed by behaviour-guiding norms that prescribe and proscribe acceptable actions and practices (Nissenbaum, 2010). For example, in the context of education, Jane’s role is a PhD candidate or student at the School of Computing and Mathematics at Charles Sturt University where some of her activities include conducting interviews for her research, analysing the data and writing up the reports. The value of the education or her role as a PhD candidate is to create new knowledge and contribute to the literature. The norms prescribe that Jane as well as other PhD candidates submit progress reports to the Research Office every six months.

Norms also define duties, prerogatives, obligation, and privileges of roles, as well as acceptable and unacceptable behaviours (Nissenbaum, 2010). Among the norms present in most contexts are those that govern the flow of information about people in the contexts. CI prescribes two types of information norms: (a) norms of appropriateness, and (b) norms of distribution.

Norms of appropriateness circumscribe the type of information about an individual that is appropriate, expected or allowable to be revealed in a particular context. For example, in the context of education, it is appropriate that the Research Office at Jane’s university know her PhD progress, but it is not appropriate that they also know details of her medical condition. Norms of distribution, on the other hand, govern the flow or distribution or dissemination of information from one party to another. For instance, within a health care context, it is appropriate that Jane’s General Practitioner (GP) should access her medical records, and divulge those details to other GPs or specialists if needed, but it is not appropriate that he provide these details to Jane’s employer. CI of the information flow is maintained when both norms are respected, otherwise, a violation of privacy occurs.

Despite the claim that CI has been proposed as a justificatory framework for evaluating new technology-based systems and practices which affect the flow of information in ethical and political terms, there is a potential criticism which goes against the framework. The critics
argue that the framework is conservative, hence, it is not easily adaptable to new technologies where there are no pre-existing expectations or norms governing the flow of information (Nissenbaum, 2010). Nissenbaum addresses the concern by responding with a decision heuristic derived from the framework and augmented by some prescriptive elements, which she notes will help explain the source of the problem and evaluate systems or practices in question. The nine steps of her decision heuristic, in brief (see more details in Section 9.3), include:

1. Describe the new practice in terms of information flows.
2. Identify the prevailing context ...
3. Identify information subjects, senders, and recipients.
4. Identify transmission principles.
5. Locate applicable entrenched informational norms and identify significant points of departure.
6. Prima facie assessment ...
7. Evaluation I ...
8. Evaluation II ...
9. On the basis of these findings, CI recommends in favour for or against the systems or practices under study ... (Nissenbaum, 2010, p. 181-183).

Nissenbaum mentions that the decision heuristic is applicable in addressing online privacy and should be done in the same way as when we analyse privacy in general because our online activities are deeply integrated into our social life. For example, banking transactions, previously rooted in our offline routine can now be done online. In order to address online privacy, one needs to locate context, explicate informational norms, identify the disruptive flows of information and evaluate those flows against norms based on general and political principles as well as context-specific purposes and values.

Nissenbaum emphasises that locating context online and explicating the norms can be performed easily for some activities which are similar to offline activities. For example, in the context of banking, whether we make a bank transaction over the phone, or face-to-face with a bank teller, or online, there is a common expectation that the transaction should be done in a confidential and secure manner. On the other hand, locating context and norms of online activities, which do not have comparable activities in our offline life, can be challenging. For instance, it can be hard to locate the context and norms of an online search engine because it might not have obvious counterparts in real life.

Consequently, Nissenbaum provides two other broad recommendations on top of the decision heuristic in order to help in locating the context and norms among online cases that do not have clear social precedents (Nissenbaum, 2011). First, we should look for the similarities between
that online case or activity and the social activity and structure. The norms or restraints on
the flow of information can be drawn out from those similarities. Second, without having to
look for the similarity of that online activity to an offline one, the norms can be established
by looking at the ends, purposes and values of the site or organisation. For instance, unlike
the context of banking, a search engine does not hold a clear social precedent. If consulting
a search engine can be seen as analogous to conducting research and seeking information in
a library catalogue, then the norms would be strict confidentiality with regard to the search
histories as practiced by many libraries.

9.3  Contextual Integrity and the tracking by SNSs

The three theories examined in Section 9.2 can be used to analyse the privacy issues that arise
from emerging technologies. Floridi’s ontological theory of informational privacy argues for
the protection of the information regardless of the zone, situation or context, because a violation
of privacy is a violation of the nature of information itself. In contrast, RALC theory is similar
to Nissenbaum’s CI theory in a sense that the zone or situation or context, not the kind of
information, is used to determined if a loss or a violation of privacy has occurred, or if the
information should be protected.

In addition to zone, situation or context, CI possesses not only two prescribed norms (norms of
appropriateness and norms of distribution), but also a step-by-step decision heuristic, which is
applicable to privacy concerns imposed by technologies. CI argues for adequate protection of
privacy for norms of specific contexts, demanding that information gathering and dissemination
be appropriate to that context. For these reasons, the researcher is interested in finding out if
tracking by SNSs is appropriate within SNS contexts. Online tracking is rather general, but
SNSs can be seen as specific contexts among other contexts online (e.g., an online shopping
and Internet banking context). Therefore, for the time being, CI is employed as a conceptual
framework for the analysis in this stage of research.

CI has been used to analyse privacy online in several contexts, for example, in the context
of public records online, and radio frequency identification (RFID) (Nissenbaum, 2004), data
mining (Nissenbaum, 2004), (Tavani, 2007b), blogosphere (Grodzinsky & Tavani, 2010), cloud
computing (Grodzinsky & Tavani, 2011) and Facebook’s News Feed feature (Hull et al., 2011).
This stage of research employs both the framework’s norms and its decision heuristic in morally
analysing the practice of tracking of users’ unintentionally shared information by SNSs.
9.3.1 Contextual Integrity’s norms and SNS tracking

According to the theory of contextual integrity, it is important to first get to know the context and to identify several variables that are involved in the information flow such as the agents (who is gathering the information, who is analysing it, and who is disseminating it and to whom), the nature of the information, and the relationships among the various parties (Nissenbaum, 2004).

Facebook, Twitter and Google are the main agents which have the ability to gather information, analyse it, and also disseminate it to various third party sites. These SNSs provide communication services to users, but at the same time, disseminate different pieces of information to the other agents such as the advertisers and the third party applications (e.g., the Farmville game application). In addition, they also have the ability to track users’ movements outside the SNSs. This is in addition to the fact that they already hold large amounts of personal information given by users. Does this mean SNSs violate users’ privacy?

Unlike other contexts like the online bookstore, or the online pharmacy, the SNS is a unique phenomena, and it is not easy to think of a comparable real life context. Looking at the SNS structure, the SNS is observed to consist of multiple profiles, and each SNS profile usually comprises of different features such as profile information, wall posts, album pictures, private messages, and advertisement banners. Every post or activity on SNS profiles are recorded and are retrievable. The notion of a community can be a useful metaphor given an SNS community has private spaces (profiles) like houses in a real community.

Let us compare an SNS profile (e.g., Facebook, Twitter, LinkedIn and Google Plus profile) to a house in a community. The process of signing up for an SNS account and getting an SNS profile is similar to the process of buying or moving into a house within a community. Each house contains different features and rooms or sections; for example, a living room area where users can hang out and interact with friends, or friends can leave messages at the door when she is not at home, photo albums, and a mailbox where private message can be delivered to the home owner.

Making one’s SNS profile private is similar to a user using a key to lock up the house to limit access from the public or random people. A public profile is like an unlocked house where anyone can just walk in. If the community owner provides different levels of privacy, the user has the ability to provide a different level of access to different people they befriend.

This comparison shows that an SNS profile can be considered as private as a house and is governed by some norms; for example, permission is needed to gain entry into one’s house. Within a community, it is appropriate that the community’s mayor conducts a census to obtain
information about the size of the population, what they do and how they live. The census data is very useful and necessary for a community leader to make informed decisions, and to support planning, administration and policy development. The census is also governed by its norms. For example, it is clearly stated in the Australian Bureau of Statistics (ABS) that any piece of identifiable information in the census data is removed and the personal information provided is kept strictly secure and confidential and is not released to other parties outside the ABS (ABS, 2012).

However, a SNS community in this case is like a small village where everyone within a community can know as much as they want about everyone else. The cookies (from both SNSs and non-SNSs) residing in a user’s computer are seen to be like many spies who would follow each resident in every step they take - either within or outside the community (e.g., to a supermarket, or a bookstore). Information gathering by the SNSs goes beyond the census norms because the SNSs not only collect information about the user (both within and outside the community boundaries), but also share user information to third party sites.

According to the norms of appropriateness, it is appropriate that SNS communities collect information about its own residents within the community, but it is not appropriate that they also collect information about the residents’s movements outside the community (e.g., Facebook is able to know that users visit taste.com.au via Facebook’s Like button). In addition, according to the norms of distribution, it is not appropriate that SNS community owners share user information, especially identifiable information, to third party organisations (e.g., LinkedIn user’s name is shared to b.scorecardresearch). Applying Nissenbaum’s contextual integrity to the context of SNS in this case study allows us to see that both norms are breached, thus, leading us to conclude that user privacy is being violated by SNSs. However, rather than relying on this analogy alone, the researcher also uses CI’s decision heuristic in order to triangulate the analysis in Section 9.3.2.

9.3.2 Contextual Integrity’s decision heuristic and SNS tracking

In this section, the researcher applies the framework’s decision heuristic to the tracking by SNSs in order to compare the analysis with the one with the prescribed norms. The decision heuristic comprises nine guiding steps detailed as follows.

Step 1: Describing the new practice in terms of information flows

SNS users provide SNSs with their information in order to sign-up for an SNS account, and later, generate other content on SNSs in order to interact with other people. SNSs are seen to be
already holding large volumes of information provided by users. On top of that, they also have
the ability to track users’ movements within and outside SNSs, particularly among websites
that use SNS widgets (e.g., Facebook’s Like button, Twitter’s Tweet button, and Google’s +1
button). They are also seen to disseminate users’ information to various third party sites (e.g.,
Facebook transfers users’ information to the Farmville application, and Twitter shares users’
clickstreams to third party sites).

**Step 2: Identify the prevailing context**

This step requires that context is established at a familiar level of generality like health care or
education. As stated earlier, unlike the context of an online bookstore, online pharmacy or on-
line banking, the SNS is a unique phenomena that does not hold a clear straightforward social
precedent, as in the case of the Google search engine illustrated by Nissenbaum (Nissenbaum,
2011). For this particular Internet service, Nissenbaum suggests that we look at its ends, pur-
poses, and values.

As mentioned in Section 1.7, boyd and Ellison (boyd & Ellison, 2007, p. 211) define SNSs
as “web-based services that allow individuals to (1) construct a public or semi-public profile
within a bounded system, (2) articulate a list of other users with whom they share a connection,
and (3) view and traverse their list of connections and those made by others within the sys-
tem.” boyd and Ellison (2007, p. 211) also proposed that what makes SNSs unique is that SNSs
enable users to articulate their social networks with people who share some offline connection
with them.

The majority of online users report signing up for an SNS account and using SNSs in order to
maintain strong ties and contacts with people they know in real life (e.g., connecting with peo-
ple they recently met or keeping in touch with friends), to express their feelings and share their
thoughts (Acquisti & Gross, 2006; Joinson, 2008; Pempek et al., 2009; Young, 2009; Cheung
et al., 2011). In addition, it has been suggested that SNS users, particularly on Facebook, tend
to search for people they know offline rather than complete strangers (Lampe et al., 2006).
Therefore, the central goal or value of an SNS is to provide a means of communication where
people can connect with one another regardless of the physical boundaries.

**Step 3: Identify information subjects, senders and recipients**

The agents, senders and recipients, refer to those from whom and to whom the information
flows, and information subject refers to the one whom the information is about (Barth, Datta,
Mitchelle, & Nissenbaum, 2006). Based on the findings of this study, SNSs (specifically,
Facebook, Twitter, and Google Plus), users and third party sites are the main agents. Users, as senders, provide or send personal information about themselves to SNSs to populate their SNS profiles and interact with other SNS users. SNSs, as recipients, collect, receive and store not only information provided by users (information subject), but also information about their browsing activities that are not voluntarily shared by users to SNSs. SNSs, as senders, transfer or leak users’ information to other recipients which are third party sites.

**Step 4: Identify transmission principles**

A transmission principle refer to the constraints on the flow of information from party to party in a context in terms of the distribution, dissemination, and transmission of information. This should detail the terms and conditions related to the appropriateness of information flow (Barth et al., 2006; Nissenbaum, 2010). The researcher examined the privacy policies of Facebook (Facebook, 2012a), Twitter (Twitter, 2012b), and Google Plus (Google, 2012b) to determine what is stated about information collection and distribution. Basically, SNSs collect any information provided and generated by their users as well as their movements or activities within SNSs, including information provided upon registration and information populated in one’s profile page. The policies also state that SNSs collect information about users whenever users visit websites or applications that use SNS platforms or widgets. In addition, SNSs also share information - both personal and non personal - about users to other parties (such as their service providers) and other users within the system.

**Step 5: Locate applicable entrenched informational norms and identify significant points of departure**

The framework suggests that established informational norms are identified, along with the significant changes brought by the practice under study. An SNS widget (e.g., a Facebook Like button, a Twitter Tweet button, or a Google +1 button) enables an SNS to collect information about their users across different sites. Facebook, Twitter and Google Plus are now able to collect and store not only personal information about their users, but also their users’ browsing activities within and outside SNS boundaries among other sites. In this case, SNS widgets resemble CCTV cameras that record users’ movements and activities. How a CCTV camera is set up can be questionable and should be investigated further as to whether or not there is a breach of privacy or entrenched norms. So now, the question is did SNSs acquire informed consent from their users regarding the use of widgets, or the collection and dissemination of users’ information?
By looking at SNS privacy policies (Facebook, 2012a; Twitter, 2012b; Google, 2012b), and regarding them as transmission principles, it appears ostensibly that they conform to what they state in their policies regarding what they collect or disseminate and how. The transmission principles also allow for commercial exchange of information in accordance with the rules of a competitive free market, but it requires the sender’s knowledge (or notice), and/or permission (consent) (Nissenbaum, 2010, p. 145).

However, it has been noted that privacy policies, and the system of notice-and-consent or informed consent, have not proved to be successful approaches to addressing privacy online (Barocas & Nissenbaum, 2009; Nissenbaum, 2011). For one thing, if we demanded all details about data collection and distribution, privacy policies would be too long, however if they are too short, they may omit vital details (Nissenbaum, 2011). For example, Google’s privacy policy alone stands at approximately 2500 words, excluding the sub links or sub pages of the policy (Google, 2012b). The other reason is that the policies are constantly changing (Barocas & Nissenbaum, 2009; Nissenbaum, 2011). Twitter has already had five versions of its privacy policy since it was launched in 2006 (Twitter, 2013). The researcher also examined the available archives of the privacy policies for Twitter and Google. Changes regarding information collection certainly exist; for example, only in the new policy (Twitter, 2012b) does Twitter address information collection via the Tweet button.

In addition, users also have difficulty in achieving meaningful understanding of the practices (information collection and distribution) to which they are expected to be able to consent (Barocas & Nissenbaum, 2009). For example, users need to have basic knowledge of technology or business in order to understand what HTTP cookies or third parties are, and in order to fully understand the following part of the privacy policy:

“...We receive data whenever you visit a game, application, or website that uses the Facebook Platform or visit a site with a Facebook feature (such as a social plugin), sometimes through cookies... Sometimes we get data from our affiliates or our advertising partners, customers and other third parties that helps us (or them) deliver ads...” (Facebook, 2012a)

Moreover, the language barrier is another obstacle faced by many users for whom English is not their mother tongue. For instance, the Facebook and Twitter privacy policies are available only in English, although recently, Facebook has made Khmer available on its site. On the other hand, Google’s privacy policy is available in many languages, although not Khmer.

In terms of choice, SNSs give users two options: either to agree with all the listed terms and conditions and sign-up for an account or to not be able to use SNS services at all. For Facebook and Twitter, by clicking on the sign up button, it automatically means that users agree to the terms and conditions. For Google, signing up will not proceed without first ticking the
check-box to agree. Users who previously signed up for an SNS account might have agreed to the terms and conditions at the time of signing up, but they may not agree to recent practice of tracking and targeting by SNSs. Facebook, Twitter and Google do however send out notifications to their users regarding the changes in their policies. But are users allowed or able to opt-out from the new policies or practices? Or while they have already provided and generated large amounts of information to SNSs, and connected with many other users, should they remove themselves from SNSs completely when they are not happy with the changes to the policies?

According to Barocas and Nissenbaum (2009, p. 5), in order for users to make an informed choice and consent, they must understand “(1) Which actors have access; (2) What information they have access to; (3) What they do or may do with this information; (4) Whether the information remains with the publisher or is directly or indirectly conveyed to third parties; and (5) What privacy policies apply to the publisher as compared to all the third parties, assuming these are even known to users...” SNSs do not make it obvious who their affiliates or advertising partners are, so users do not have access to the privacy policies of those third parties regarding the use of their information. In addition, because of the length and changeability of privacy policies, as well as jargon and language barriers within the policy contents, we can assume that SNS users do not fully understand or are not aware of (1) Who the stakeholders are, (2) What information about them is gathered by SNSs and third parties, (3) How their information will be used, (4) Whether their information remains with SNSs or shared to third parties, and (5) What policies apply to SNSs, third party sites, as well as any websites they have visited or wish to visit. Therefore, Facebook, Twitter and Google Plus did not gain informed or active consent from their users in moving their information across different contexts and in tracking their users’ online movements. To be able to give informed consent, users are expected to read carefully to fully understand the privacy policies of websites they intend to visit.

If we rely solely on SNSs’ privacy policies to define the norms in the information flows within SNS contexts, it appears that SNSs conform to what they state in their policies regarding the collection and dissemination of users’ information. However, based on the arguments and rationales regarding the inadequacy of privacy policies, it appears that users are not explicitly informed about the tracking by SNS widgets, and moreover that SNSs did not gain their informed consent in setting up and using these widgets. The departure from the entrenched practice in this context is the SNS widgets that resembles hidden CCTV camera used to capture users’ movements and activities across different sites. Prior to the existence of these widgets, SNSs can only record users’ activities within SNS boundaries.
Step 6: Prima facie assessment

Within this step, a decision can be made as to whether or not the practice under study violates the information norms. By proceeding through Steps 1 to 5 of the decision heuristic and by examining privacy policies among SNSs (Facebook, Twitter, and Google Plus), it is argued that SNSs did not gain informed consent from their users regarding the collection and distribution of their information. SNS widgets are seen to be hidden CCTV cameras that are used to record users’ movement within and outside SNS boundaries. The descriptive component of the heuristic establishes that such tracking constitutes a breach of information norms, and consequently a violation of contextual integrity. However, further steps assessing the legitimacy of the tracking are required by the heuristic.

Step 7: Evaluation I

The framework also suggests that we consider the moral factors affected by the practice in terms of harms and threats to autonomy and freedom. Regarding privacy policies as transmission principles, Facebook, Twitter and Google Plus did not gain informed consent from their users in recording their online movements across different sites and in moving their information across different contexts. Users are not aware of this tracking or recording; hence, they have the impression or perception that their online activities are private.

On the bright side, this tracking allows SNSs to provide relevant advertising contents to their users. The ads are usually based on their users’ browsing preferences or habits. For example, by visiting a recipe Web site like Taste.com.au, the theme of cooking can be drawn from this browsing; and relevant advertisements for users could be sales or bargains related to kitchen utensils or cookbooks.

On the other hand, the tracking or recording by the hidden camera appears to be doing something wrong and affects one’s autonomy. Personal autonomy is one of the four functions of privacy that Westin (1966) refers to as “the desire to avoid being manipulated or dominated wholly by others.” Westin argued that privacy protects autonomy in terms of individuality and consciousness of individual choice in life. He also added that another aspect of autonomy is the individual’s sense of “he” who decides when to “go public”. In this case, SNS users are being observed and their browsing activities within and across different sites are recorded by SNSs. Therefore, online users’ autonomy is deteriorated to some extent because users did not decide or choose to share this type of information with SNSs.

Without the fear of being observed, or judged by others and without the constraints and expectation of a convention, an individual has more freedom to formulate life choices, preferences
and commitment for themselves (Nissenbaum, 2004). The context of SNS tracking without users being well-informed impinges on users’ freedom to a degree. Users’ choices of browsing activities could be different if they were aware of tracking or monitoring, or if they knew the Internet is not a private space. In this case, little do they know that SNS widgets can serve as a recording device, nor are they well informed about the monitoring by SNSs. Hence, they have an expectation that their online activities are private, and they believed they had complete freedom in a monitored or observed zone. Their online activities or behaviours could be different if they were aware of SNS tracking.

To sum up, users thought they had complete freedom and autonomy regarding their browsing activities, and presumably, they thought their online activities were in a **private zone**, when in fact SNSs are able to know what they are doing online.

**Step 8: Evaluation II**

In addition to considering the impact on autonomy and freedom, the framework also suggests that we consider whether the practice also impinges on the values, goals, and ends of the context.

As outlined earlier in Step 2 of the heuristic, the central goal or value of SNSs is to enable communication between users who may be in different physical locations. The majority of users are active on SNSs for the purpose of communication, enabled by SNSs. The goal and mission of Facebook is to make the world more open and connected by giving people the power to share (Facebook, 2012b). Facebook is seen to enjoy tremendous success while attracting more than a billion active users owing to the fact that they succeed in delivering such a communication service **expected** by their users.

SNS users are usually allowed to set their profiles to a private mode, where they can restrict visibility to a limited audience. However, the notion of “private profile” gives users an inaccurate impression that only their limited audience would see what they share because most users have no idea that their information and browsing activities are being shared to third party sites, and tracked by SNSs in the HTTP headers.

Facebook, Twitter and Google Plus do deliver the expected communication services based on their goals and mission. However, the SNS tracking appears to go beyond their mission as well as users’ expectations of connecting and communicating with their friends. Within the context of SNSs, users expect to be able to share experiences and communicate with other users, but not to be tracked and monitored across different sites. Privacy policies which are frequently lengthy and changeable, fail to formally inform users about tracking. Therefore, users think or
expect that their online activities are private, when actually they are not.

**Step 9: Recommendation**

This is the stage whereby the framework recommends in favour of or against the practice under study on the basis of the findings of the previous eight steps (Nissenbaum, 2010).

Earlier, the researcher looked at the ends, purposes and values of SNSs and examined the privacy policies of Facebook, Twitter and Google Plus in order to locate the entrenched norms. It was found that those SNSs already listed all terms and conditions of their current practice, including data collection via the use of widgets. However, after examining the approach to online privacy by analysing privacy policies and their inadequacy, it is argued that SNSs fail to gain informed consent from their users in collecting and disseminating their information. SNS widgets (Facebook’s Like button, Twitter’s Tweet button, and Google’s +1 button) are seen to be hidden CCTV cameras or recording devices that can monitor users’ online activities across different sites.

In Step 7 it was argued that users’ autonomy and freedom are being impinged due to the fact that users are not formally or explicitly informed about the monitoring, and nor did they make the decision to share their browsing information with SNSs; therefore, they assume what they are browsing is private when in actual fact it is not. Meanwhile, in Step 8, it is also argued that the practice additionally impinges on the context’s goals, values and ends owing to the fact that within an SNS context, users have expectations of being able to connect and communicate with other users, but they might not expect that their online activities across other different sites can also be monitored by the SNSs.

Does tracking by SNSs really improve their users’ experience? Prior to the availability of the SNS widgets, users simply copied a specific URL from a page and pasted it in their profile page in order to share that link with their friends. SNSs were able to know what users did only within SNS boundaries. What SNS widgets enable is a single click (either on Facebook’s Like button, or Twitter’s Tweet button, or Google’s +1 button) that allows users to share things on their profiles.

On one hand, the SNS widgets improve their users’ experiences in sharing their preferences while also enabling the SNSs to track their users’ movements across different sites that embedded those widgets; therefore, enabling SNSs to provide better targeted advertisements to users. On the other hand, with the absence of SNS widgets, users can still share, and hidden CCTV cameras are not present. The advertisements will still be there except they are not as personalised as when the widgets were present.
The researcher also examined privacy settings of Facebook, Twitter, and Google Plus. Currently, Facebook and Twitter provide users with the ability to manage the widgets. Managing Facebook’s widget means users have the choice of whether or not to display the pages they “Liked” on their profiles or newsfeed, as shown in Figure 9.1. Meanwhile, disabling Twitter’s widget means “Stop Twitter from tailoring Twitter based on my recent website visits”. Google Plus does not provide users with the option of disabling or enabling their widget. Nonetheless, managing Facebook or Twitter’s widget only limits the visibility of tailored advertisements or Tweets and browsing activities from users and their friends, but does not stop Facebook and Twitter from tracking online movements, because their widgets are still present after they were disabled via Facebook and Twitter privacy settings.

Therefore, based on the rationales and arguments so far, the framework recommends that we reject the practice of tracking and monitoring users by SNSs, like Facebook, Twitter, and Google, across different sites. Within an SNS context, it is appropriate that an SNS collects users’ information within its boundary for the purpose of improving users’ experiences in connecting with other users, and for the purpose of improving services in terms of efficiency in troubleshooting and analysis of the site performance or network traffic. But it is not appropriate that an SNS also collects information about its users’ online movements outside an SNS boundary, and disseminates their information to other parties without users being explicitly informed.

Figure 9.1: Managing Facebook’s widget
9.4 Summary

The CI framework, introduced by Helen Nissenbaum, is intended as a framework to analyse people’s reactions to a technology or system that affects information flows, as well as a tool to evaluate the practice of a technology or a system from a moral viewpoint. The framework focuses on a context and the flow of information, rather than the types of information. It also argues for protection of information in the public space and any non-sensitive information divulged in a specific context. The fact that information is infused or available in public or information is non-personal or non-sensitive does not mean it can be up for grabs. In addition, Nissenbaum also developed a decision heuristic (derived from contextual integrity) which assists the response to privacy challenges posed by new emerging technologies like SNSs. The decision heuristic suggests that we locate contexts, explicate informational norms, identify the disruptive flows of information, and evaluate the flows against the norms based on general moral and political principles as well as context purposes, ends, and values.

This stage of the research focuses on the application of Nissenbaum’s CI framework (its norms and decision heuristic) to the practice of online tracking by SNSs. This tracking is enabled by the use of SNS widgets by first party sites, and was discovered in the exploratory study and experiments among 20 participants. In addition, SNSs are also sharing users’ information to third party sites who appear to be their advertising partners. By using Nissenbaum’s contextual integrity and by looking for the context and the variables involved in the information flow, SNSs like Facebook, Twitter, and Google Plus are seen to be the main agents with the ability to gather, analyse, and disseminate users’ information to third party sites.

The comparison of an SNS to a real life community, specifically a user profile to a house, helps to identify some norms which may govern a community and data collection within that community context. According to the norms of appropriateness, it is appropriate that SNSs collect information about their users within the site, but it is not appropriate that they also collect information about their users outside SNS boundaries. Meanwhile, according to the norms of distribution, it is appropriate that SNSs use information collected to assist in maintaining the site traffic and performance, and improve users’ experiences in the SNS, but it is not appropriate that they also share this type of information with third party sites. The analysis based on contextual integrity’s norms revealed that both norms are breached, leading to the conclusion that users’ privacy is being violated by SNSs.

In addition to the framework’s norms, the researcher also triangulated the analysis by using the decision heuristic derived from the framework. The situation was analysed by first locating the ends, values and purposes of SNSs, and by going through the nine steps prescribed by the decision heuristic. The researcher also examined the privacy policies of Facebook, Twitter, and
Steps 1 to 4 of the heuristic introduced us to what the practice is, who the stakeholders are, what the context is and what the transmission principles are. Users, SNSs and third party sites were the main agents or stakeholders while users’ information flows across different contexts between first party and third party sites. In Step 5 of the heuristic, it was discovered that there was a breach of the informational norm from the practice under examination when SNSs failed to gain informed consent from their users because of the weaknesses of privacy policies. SNS widgets are seen to be analogous to hidden CCTV cameras used to record and monitor online users’ movements across different sites.

In Steps 7 and 8 it was argued that users’ autonomy is being threatened because they are not explicitly informed about tracking, and tracking also impinges on the goals, values and ends of an SNS where users expect to be connected with other people, but not to be tracked and monitored by an SNS. Meanwhile, in Step 9 of the heuristic, the framework proposed that the practice is rejected by the framework because online users’ right to privacy is being violated by the practice of SNS tracking across different sites.

However, we live in a society where technologies are evolving and always moving forward. There should be a trade off or a balance of benefits of the widgets to both users and SNSs. Hull et al. (2011) stated that it was the design of a system, in their case, Facebook’s News Feed feature, that may raise privacy issues, although this can be addressed through a better design so that information flow can be more transparent to users. Likewise, the practice of using SNS widgets which leads to tracking of users’ movement across sites, can be favoured or accepted by the CI framework if there is a change in the design in which SNSs give users choice or freedom to enable or disable the widgets (or hidden camera) completely. Or at least, the design should make this tracking explicitly clear to users so that they are well-informed about such a practice and are able to make better decisions with regard to using SNS services or browsing online.
Chapter 10

Discussion

In this chapter, the intention is to discuss the outcomes of this research thesis in light of the research questions and existing literature. It will also discuss the implications and make recommendations about the current issues of online tracking by SNSs. The chapter is organised as follows, with Section 10.1 recapping the research questions, while Section 10.2 summarises the findings. Meanwhile, the research findings as well as recommendations are discussed in Section 10.3 and Section 10.4, respectively.

10.1 Restatement of research questions

The research began with the researcher browsing online while having the HTTP traffic resulting from these browsing activities captured and analysed. Among other outcomes, the results suggested that her unintentionally shared information were being collected and tracked by numerous third party sites such as advertisers, data aggregators and SNSs like Facebook, Twitter and Google Plus. This tracking shaped the research focus and questions in that privacy surrounding this practice by SNSs became the target to be investigated from the perspective of Cambodian online users and from the perspective of the CI framework. The questions that guided this research were:

1. What is the nature of information sharing in HTTP headers?
   (a) What type of information is being shared?
   (b) With whom is this information being shared?
   (c) What are the implications of this sharing?
2. What are SNS users’ views on online tracking and privacy?
   (a) What does privacy mean from SNS users’ perspectives?
(b) To what extent are SNS users aware of online tracking in general as well as online tracking by SNSs?
(c) What are SNS users’ reactions to the practice of sharing and tracking information online?

3. What are the ethical implications of the collecting and tracking of users’ information by SNSs?

10.2 The results

10.2.1 Unintentionally shared information in HTTP headers

The first stage of this research aimed to address the first research question and its subsidiaries. It focused on the gathering and sharing of users’ information in HTTP headers, particularly their unintentionally shared information. In Chapter 4 the researcher conducted an exploratory study in which she browsed online and analysed HTTP headers or messages resulting from her browsing activities. In order to triangulate the results as well as assess the feasibility of larger scale experiments, the exploratory study’s experiments were conducted on computer devices running different operating systems such as Mac OS, Windows and Linux, with and without proxy servers.

The Wireshark experiments with the 20 Cambodian participants allowed the researcher to investigate the implications of information leakages among a diversity of browsing behaviours and devices. The researcher first conducted a workshop where all participants were introduced to the research details and intention as well as Wireshark and some basic knowledge of HTTP cookies. Participants were asked to follow a set of protocols. For example, they were required to clear the cookies and browsing histories in their browsers before running Wireshark. All of them chose their own sites and online activities to be performed for 10 minutes while having them recorded by Wireshark. Then the researcher collected Wireshark files from each participant, and analysed and revealed the findings to participants during the open-ended structured interviews.

The findings from both the exploratory study and Wireshark experiments with 20 participants indicated that users’ unintentionally shared information can be gathered and monitored by numerous third party sites that are normally unknown to online users. Third party sites in this study are found to be advertisers (e.g., Doubleclick, Google, Yahoo), data aggregators (e.g., Scorecard Research), and SNSs (e.g., Facebook, Twitter, and Google Plus). The researcher’s and participants’ unintentionally shared information was observed to propagate to more than
one level of third party sites. It is transferred from first party sites to third party sites, and from these third party sites to other third party sites, up to 5 levels of propagation. In addition to the extraneous connections from the visit to first party sites, participants’ browser toolbars were also seen to initiate connections to third party sites like Google and Yahoo.

Connections to third party sites by first party sites allowed those third party sites to track users’ online movements. Even though participants’ toolbars did not carry their personal information, the connection to the same third party sites by the toolbars and the first party sites enabled these third party sites to identify a particular user and track their online movements across different sites. SNSs like Google, Facebook, and Twitter are able to acquire not only participants’ personal information provided by them, but also participants’ browsing behaviours across different sites. Hence, they are problematic for users because those SNSs are able to track users’ online movements and can link those details to their profiles. For these reasons, the rest of the research focused on the tracking by SNSs, from the perspectives of SNS users and from the perspective of the CI framework. The research then proceeded to determine whether the research participants were aware of this tracking and how they perceived privacy online.

10.2.2 Online tracking from the perspective of Cambodian participants

The second stage of the research aims to address the second research question and its subsidiaries. It gauged Cambodian research participants’ perception and awareness regarding privacy and online tracking, particularly the tracking by major SNSs, based on their online browsing experiments. This stage of the thesis is a qualitative study. It is influenced by the ethnographic method and employs interviews as a data collection technique. Ethnographic method seeks to understand people’s views and opinions, while open-ended structured interviews, in which participants were asked the same series of questions, allows participants to express their thoughts.

From the perspectives of Cambodian research participants, the meaning of privacy revolves around the notion of an individual’s right, and/or control over others’ limited access to their information. Relationship with others determines whether certain types of information are private. For example, family matters which are considered private by the majority of participants can be shared and discussed with their close friends, but not with the public. They do not disclose or discuss these matters on SNS walls that can be seen by other users.

The findings revealed that, while signing up and using SNSs on a regular basis, participants in the study revealed a large amount of factual information (e.g., name, e-mail address, picture and date of birth), and information about their interests (e.g., favourite movies, and books) because
they thought they were required to do so. They also generate and share other information while engaging with other SNS friends because they enjoy sharing and they like to seek emotional support. This suggested that participants in this study intentionally disclosed a large volume of personal information that is then stored by SNSs.

Regarding their awareness of online privacy and online tracking, the majority of participants did not read privacy policies. These policies were perceived to be too long and difficult for them to understand and even if they read them, they had no choice but to agree with all the terms and conditions in order to use the services provided by the sites. Almost all participants (95%) were not aware of the online tracking by the third party sites or advertisers. They understood online tracking to be tracking by other Internet users who were able to glean information they posted on SNSs.

After the researcher revealed that their online movements were being tracked by third party sites like advertisers as well as SNSs to whom they voluntarily provided their personal information, all participants were shocked. They became angry and sad and expressed concern about the loss of privacy and the potential for judgement by others based on their online activities. Nineteen of the 20 participants were not aware of any tools or technologies available to reduce the tracking by third party sites. Only one participant reported using the Firefox extension intended to reduce the number of advertisements on the Web page.

Findings from the investigation of information sharing in the HTTP headers and from the interviews show that there is a gap between the practice (of sharing information to third party sites and tracking) and Cambodian participants’ awareness of the practice. These findings suggest that Cambodian participants do care about privacy based on their perception of privacy. However, they lack of technical knowledge in understanding how their unintentionally shared information can be tracked and recorded. They also expressed concerns about this type of tracking by SNSs. The findings also suggest SNSs are storing large volumes of both participants’ intentionally and unintentionally shared information. Next, the researcher used Contextual Integrity, which is a privacy framework, to analyse the current practice of SNSs sharing and tracking user information from a moral viewpoint.

10.2.3 Tracking by SNSs from a moral perspective

Tracking by SNSs is enabled by the use of SNS widgets (e.g., Facebook’s Like button, Twitter’s Tweet button and Google’s +1 button) among the first party sites visited by the researchers and participants in the study. In Chapter 9, such a practice was analysed using Contextual Integrity as a framework.
Helen Nissenbaum’s Contextual Integrity (CI) accounts for people’s reaction to a technology or a system that affects information flows while also helping to evaluate the practice of a technology or a system from a moral viewpoint. *Norms of appropriateness* of CI determine the types of information about an individual that is allowed to be disclosed in a particular context, while *norms of distribution* is concerned with the flow or distribution of information from one party to another.

The use of the framework’s norms - norms of appropriateness and norms of distribution - helps to identify the norms which may govern the information flow within the system. It is appropriate that SNSs collect information about their users within the sites, and use those types of information to maintain the site traffic and performance while also improving SNS users’ experience. However, it is not appropriate that SNSs collect information about their users outside SNS boundaries while also sharing their users’ information to third party sites. Within the SNS context, the flow of information was disrupted.

In addition to using the framework’s norms, the study also employed Nissenbaum’s decision heuristic which is derived from CI. The decision heuristic suggests that the context is located, the norms are explicated, and the disrupted flows of the information are identified. While proceeding through Step 1 to Step 9 of the heuristic, a breach of information norm is discovered in Step 5 due to the fact that SNSs failed to gain their users’ informed consent in tracking their online movements. While an SNS can be compared to a real life community, SNS widgets in this case can be compared to *hidden CCTV* cameras that are used to capture people’s movements, not just within but also outside the community.

Users’ autonomy was seen to be threatened by such a practice in Step 7 and 8, because users were not explicitly informed about the tracking. Hence, they did not have the autonomy to decide whether they wanted to share information about their browsing to SNSs. The practice also impinged on the goals, values and ends of SNSs where SNS users expected to be able to communicate with their friends, but instead, they are also being tracked by SNSs. Step 9 of the heuristic suggests a rejection of the practice because it violates users’ right to privacy.

### 10.3 Discussion

The main focus of the first stage of this research is the tracking of users’ unintentionally shared information by major SNSs. Previous works were concerned about the tracking by third party sites like advertisers and data aggregators. They were also concerned that users’ personal information was being leaked from SNSs (e.g., Facebook, Twitter, and LinkedIn) (Krishnamurthy & Wills, 2006a, 2006b; Krishnamurthy et al., 2008; Krishnamurthy & Wills, 2010a, 2010b;
Mayer, 2011). However, SNSs’ widgets subsequently emerged which enable SNSs to track users’ online movements across different sites that embedded those widgets (e.g., Facebook’s Like button, Twitter’s Tweet button, and Google’s +1 button).

As mentioned earlier in this thesis, this type of tracking is relatively new and becomes a concern because SNSs are able to acquire not only users’ intentionally shared information (e.g., name, pictures, contact details, location, and other activities within SNSs), but also their unintentionally shared information (e.g., browsing activities among sites other than SNSs). The combination of these two types of information revealed much about a person’s life. For these reasons, the latter two stages of this research also focused on the tracking of users’ unintentionally shared information by SNSs, from the perspectives of online users and a privacy framework.

In addition, previous literature did not account for online users’ awareness regarding the tracking by SNSs, nor did they focus on participants or online users from Cambodia. This research covered not only the technical aspects, but also social aspects by having Cambodian research participants taking part in Wireshark experiments, then in qualitative interviews. The interview questions and focus of research are shaped based on the findings from the researcher’s and participants’ browsing activities. Furthermore, this is the first ICT study that involved Internet users from an under-researched country in terms of ICT areas like Cambodia.

Similarly to previous works (Young, 2009; McDonald & Cranor, 2010a, 2010b; Al-Saggaf, 2011; Wills & Zeljkovic, 2011), the findings in this research suggested that Cambodian online users do care about privacy. This is ascertained through their definition of privacy (as described in Section 10.2.2) and through the fact that they did change their privacy settings in SNSs to limit their audience. In addition, they had negative reactions toward the fact that their online activities can be tracked and recorded by major SNSs. However, the level of awareness of online tracking among research participants from Cambodia appears to be lower than that of other countries such as the United States (TRUSTe, 2008; Wills & Zeljkovic, 2011). This may be due to the fact that these research participants do not have a substantial background in ICT, or that Cambodian online users lack knowledge in understanding the technical aspects of online tracking. Future work may also assist in assessing the awareness of online tracking among participants who have a background in ICT.

In addition to looking at the tracking by SNSs from the perspective of SNS users, this research also employed the CI framework in order to analyse such a practice from a moral perspective. This has not yet been covered in the literature. As summarised in Section 10.2.3, the analysis suggested that the practice of tracking users’ unintentionally shared information without users’ informed consent violates users’ right to privacy. This outcome appears to be consistent with
the reactions of participants in the second stage of the research, where they expressed concern about privacy loss as a result of this practice. This is the very first work that not only investigated the information sharing in the HTTP headers, but also analysed the findings from the perspectives of online users as well as a moral framework.

Although this research focused mainly on concerns in regard to the tracking of users’ unintentionally shared information by SNSs, it discovered other results that were not covered in the literature. For example, this research discovered that users’ unintentionally shared information traversed more than just one level of third party sites. It propagated from visited or first party sites to third party sites, and from these third party sites to other third party sites, as shown in Figure 10.1. Those third party sites, which do not have any affiliation with online users, are able to learn about users’ online activities. Future works can delve more deeply into this propagation in order to find out how much information is transferred to those third parties, and also to understand the affiliation or business agreements among first and third party sites.

Figure 10.1: The flow of an online user’s unintentionally shared information

In addition, the Wireshark experiments among research participants in Cambodia on different personal computers discovered that participants’ browser toolbars also sent out multiple HTTP connections to third party sites like Google. This type of HTTP connection was not covered in previous works that focused on information flow in the HTTP headers. This finding suggests that other extraneous contents on the browsers (e.g., browsers’ add-ons, and toolbars) also play a role in enabling the tracking of users’ online activities by third party sites. Future works may focus on this type of HTTP connection in order to better suggest appropriate strategies that help to reduce the chance of tracking.

To sum up, these research findings indicate that online activities or intentionally shared information can be tracked by numerous third party sites including SNSs like Facebook, Twitter and Google Plus. The study also suggests that there is a gap or disconnection between the practice of information tracking and Cambodian participants’ awareness: 19/20 of them are not aware of online tracking at all and they expressed concerns about loss of privacy due to this practice. The results are not generalisable to the whole population of online users in Cambodia, but it shows
that more and more users enjoy using the so-called free services online while generating and giving up large volumes of information about themselves to numerous unknown or unknowable parties. The issue of privacy is becoming more important because of the technology’s ability to collect, store, and discover new knowledge or make inferences from the users’ information collected online. The analysis of the practice from a philosophical viewpoint by using the CI framework suggests that the practice violates online users’ privacy because SNSs collect users’ information across different sites without obtaining their informed consent.

10.4 Is there any hope after all?

The findings of this research suggest that online tracking is still happening and what is new is the tracking by SNSs. Despite all the efforts from researchers in different areas (e.g., technical, social, and philosophical) in bringing this matter into the spotlight, online users, particularly research participants from Cambodia, have little knowledge of the practice. Users are one of stakeholders involved in this practice, and they are also the customers or targets of the businesses (e.g., targeted advertisements and sales). Associated risks to privacy are not widely perceived by online users either. Users enjoy using the so-called free services offered online including e-mail and SNS services, while also sharing a large volume of information about themselves and communicating with people they know online and offline.

As identified within the findings of this research, there is a gap between the practice of online tracking and users’ perception of this practice. The following scenarios are intended to provide one of the possible reasons why the gap exists. They also help to illustrate the consequences of online tracking and to show why privacy is significant and should not be neglected, while also showing the difference in information flow in different contexts or situations.

Scenario 1: Soriya and her shopping activities at a local market in Cambodia.

In Cambodia, supermarkets are not as popular as local markets where farmers sell their fresh produce from different places. Khmer recipes are known to consist of fresh vegetables and fruits. Soriya likes to shop for groceries at her local market. She regularly gets fresh ingredients (such as fish, lemongrass, mushrooms, eggplants, etc.) from a number of particular farmers; hence, becomes their regular customer. When she regularly returns to the same sellers, the sellers know what she likes or wants, and that if what she wants is not available, they may be able to recommend suitable substitutes.

In this scenario, Soriya does her shopping at a local market - a public space. Her activities can be seen by many people. The local farmers or sellers with whom she interacts may re-
member what she likes or what she usually buys from them, but they do not have any written or recorded information about her, her credit card or postal or residential address. The stakeholders involved in this case include Soriya, the agent from whom the information flows, the farmers to whom her interest or preference is known, and other strangers who may have noticed Soriya’s movements in the surrounding neighbourhood.

**Scenario 2:** Soriya and her shopping activities at the supermarkets in Australia.

Supermarkets are more common in Australia than Cambodia. There are items that Soriya can purchase at the supermarkets apart from just groceries. The service she gets while shopping at a local market in Cambodia is different to the one offered by the supermarkets in Australia because there are too many customers for the supermarkets to get to know them personally and individually. However, by being a loyal customer to a supermarket and by possessing and using the supermarket’s reward card, Soriya usually gets discounts on her shopping. In this case, the supermarket is able to record information about her and her shopping such as what she usually purchases and how much she spends.

In this scenario, supermarkets are able to record and store information about Soriya and her shopping details because she registered to be their loyal customer in exchange for the rewards such as discounts or special promotions. Her information, as well as other customers’ information can be used by the supermarkets for the purpose of data mining where new knowledge or patterns about customers’ shopping can be revealed. However, if she chooses not to become a loyal customer to any supermarket and if she only purchases and pays for her shopping by cash, the supermarkets are not able to acquire information about her. Other customers or Soriya’s friends may see her shop at a particular store, but there will not be any written or recorded information about her and her shopping in any database.

**Scenario 3:** Soriya and her online activities including online shopping

Soriya also enjoys shopping online which is not available in Cambodia. A variety of items can be purchased online. For example, she likes to order her swimming gear and other accessories from Speedo, while she usually browses and buys skin care products from eBay. She also manages her banking transactions online. Like many other Internet users, she regularly uses Facebook, e-mail services and reads news articles online. Unlike shopping at a local market in Cambodia, her online activities can be recorded by both first and third party sites. Those sites can target her with advertisements based on her past and present interests.

When Soriya browses online as described in Scenario 3, information about her, including both identifiable and non-identifiable information, propagates to numerous parties including first and third party sites. As discovered in this research, online users’ activities are being tracked and recorded by numerous third party sites such as advertisers or data aggregators as well as
SNSs. eBay is able to learn and record Soriya’s favourite products for skin and hair care and this information can be shared or sold to other parties that are not known by her.

In those three scenarios, Soriya shops in the local market, she interacts with others including the sellers or people she may know who shop in the same area. Hence, she may have a prior assumption that her activities in the market are not completely private. Meanwhile, when Soriya shops in the supermarket and when she uses her loyalty card, she may also be aware that her shopping details are recorded by the supermarkets, or at least are seen by the checkout shop assistant. So Soriya does not expect complete privacy in her activities in the supermarket. However, the tracking and recording activities by those sites are not visible to Soriya or other Internet users. They do not interact with third party sites, nor with SNS widgets. Participants from Cambodia assume they have privacy if they use the Internet in their locked room without anyone seeing what they do online. However, they may not have privacy if they post information on SNSs like Facebook, because Facebook users or friends are able to see their posts.

This suggests that users tend to be carefree about online tracking because the tracking done online or automated by various agents (including first and third party sites) is not visible or obvious to them. If the tracking occurs in real life, they may react differently. When Soriya or other customers signed up for a reward card from a supermarket, they may be happy to let supermarket keep track of their shopping with one swipe or one scan. However, they could be very uncomfortable if they went to the same supermarket, and at the checkout, they had to report information about themselves (full name, address and bank details) and their shopping details (what they bought and how much they spent) to a shop assistant who take notes and writes down details about loyal customers.

Similarly, users may feel comfortable when browsing online alone in a room, or even in a public space like the library where no one pays attention to what they do on their computers. However, it seems very unusual to have someone or a group of people watching over their shoulders while they use their computers to go online. Likewise, it does not feel right having people walking behind us and taking note of which store we visit and what we purchase in a shopping mall. Online tracking occurring without users’ awareness can be seen as users not being aware of people watching over their shoulders or following them from store to store, whereas the online tracking that obtained users’ consent is more transparent, but users cannot refuse the practice.

In the next section, a number of options will be discussed that may shed light on the current practice of tracking and users’ concerns about privacy loss.
10.4.1 Browser level protection

The first aspect to be discussed is protection at the browser level, including the use of browser settings and browser extensions. Online users can clear cookies and browsing histories, and disable the script execution through browser settings. However, disabling JavaScript execution can negatively affect page quality (Wills, 2011). In addition, existing extensions such as AdBlock (Palant, 2013) and Ghostery (Ghostery, 2013) can be used to remove or block extraneous contents such as advertisements on the first party site pages, whereas Priv3 extension can protect users from tracking by SNSs if users do not interact with those sites (Priv3, 2012). Unfortunately, these tools did not prove to be an effective solution to the problem because, firstly, they did not completely eliminate the tracking (particularly by SNSs) and secondly, not many users were aware of online tracking (Krishnamurthy & Wills, 2006a). In addition, they also require installation by users, which can be daunting for some users.

Meanwhile, in the context of information security, the browsers show warnings to users when an attack might be occurring; then, users can decide to either navigate away from the sites or click through the warnings. Akhawe and Felt (2013) found that those browser security warnings can be effective in practice as they enable the system to communicate security to end users. In addition, in a separate study by Braunstein, Granka, and Staddon (2011) it has been found that privacy wording have a strong impact on user behaviour: priming users to think about privacy heightens their attention to the potential privacy risks. These show that user experience of a warning can have a significant impact on their behaviour and awareness. Similar concept can be used in the context of privacy and online tracking. For example, the browser may communicate privacy to end users when online tracking by third party sites is occurring. Then users can choose to continue their visit or leave the page. This eventually raises user awareness in regard to the third party site tracking.

10.4.2 User awareness and informed consent

The second aspect to be considered is user awareness and informed consent as information collected is about online users. As discussed in Chapter 9, tracking by SNSs violates a user’s right to privacy as SNSs failed to get users’ informed consent. In order to make an informed choice, users must know all the details involved in the information collection such as who has access to what information about them and how their information is processed or used once collected (Barocas & Nissenbaum, 2009; Krishnamurthy, 2013).

In order to provide truly informed consent, a user would probably need to know exactly what is described in the terms and conditions or privacy policies. In the current privacy policies,
there are not many details about who exactly the third parties are or to what extent the information will be used, or how long it will be stored. In Chapter 9, the researcher discussed the ineffectiveness of the privacy policies. For example, if everything is included, the policies will be too long. They also contain jargon that general users may not be able to fully understand. Trying to give too many details in a very long privacy policy may not be a good idea - it only confuses and frustrates users (FTC Staff, 2010). The language could be a barrier for users who are from different backgrounds and cultures, and who speak languages other than English. Therefore, getting informed consent from online users worldwide and raising awareness about online tracking and precaution could be hard work.

In the health care context, when a patient is admitted to the hospital for surgery, she needs to sign a consent form that informs her about the procedures the doctor will undertake. The patient does not need to have great knowledge in medical terms in order to understand the procedures, but to the extent that she is aware that something in her body is not well. The procedures of information collection could be simplified so that general online users who do not have a deep understanding of the technology can still make sense of such a practice. For example, users may not need to have a great knowledge of Computer Science in order to understand the practice, but it would be sufficient if they know what information about them is being collected, by whom, and how.

As discussed in the shopping scenarios earlier, online tracking is like having people walking behind shoppers taking notes about every visited stores, and purchased items. Having informed consent in this scenario is like the shoppers being aware and informed that if they enter this mall, there will be others walking behind them, yet they have no choice but to agree to this. There are users who may be enjoying targeted advertisements and some who prefer to have private browsing (McDonald & Cranor, 2010a; Wills & Zeljkovic, 2011). Having a choice to be or not to be tracked by third party sites or SNSs would make the situation different.

### 10.4.3 User choices and decisions

Another aspect to be considered is SNS users’ choices or decisions to either opt in, or out of the tracking by SNSs. For example, in the scenario where Soriya does her shopping in the supermarket and chooses to pay by cash rather than with her bank card, and she is not a loyal shopper at any supermarket, there is no written record about her in the supermarket’s database. In the same way, there could be changes in the current system designs that give users choices and options. For example, users who like to get targeted advertisements or promotions can choose to be tracked, while users who like to have private browsing should be able to not be tracked.
Paying to have a private SNS profile and private browsing could also be another option because currently it appears that users are [unknowingly] trading their information for the so-called “free services” offered by SNSs. A similar concept is widely adopted among mobile applications such as those in the Apple Store. For example, there are two versions of the “4 pics 1 word” application: a free and a paid version. A free version application is usually loaded with ads while the paid one is ad-free. However, further work is needed to assess the effectiveness of this option.

10.4.4 SNSs

A further aspect to be discussed is the involvement of related SNSs such as Facebook, Twitter and Google. As considered in Section 10.4.3, some changes in the system design may bring changes to the current situation of online tracking. SNSs hold the power to make changes because they have control of the systems, and they understand the economic value of online tracking (Krishnamurthy, 2010; Gill et al., 2013). SNS users are SNSs’ customers; users’ information is the main income generator for SNSs. Hence, keeping the customers happy would secure business sustainability for SNSs.

Companies should make consumer privacy a basic consideration and promote consumer privacy throughout their organisations at every stage of their product or service developments (FTC Staff, 2010). In addition to incorporating substantive privacy protections (e.g., data security, collection limits, retention practices, and data accuracy), companies should also incorporate computer ethics, which is the analysis of the nature and social impact of computer technology (Moor, 1985), into their culture. By doing so, they are able to perceive the social and legal consequences of their actions, products, or services before making any decision (Sherratt, Rogerson, & Fairweather, 2005).

It is also crucial that SNSs increase the transparency of their data practices in such a way that SNS users feel comfortable: what data is collected and how it is used (Awad & Krishnan, 2006; FTC Staff, 2010; Whitten et al., 2011). For example, SNSs should make their privacy policies or notices clearer, more standardised to enable better comprehension of the data collection, particularly from users whose first language is not English, like Cambodian SNS users. In addition to transparency, SNSs should also allow users to opt out of the tracking if they are not comfortable with it, while still providing free services to them. Giving choices to users will not only help users make decisions during the collection of their information, but also will facilitate their ability to compare privacy options offered by different SNSs. This will eventually promote competition on privacy (FTC Staff, 2010).

Unfortunately, allowing users to opt out of the tracking could result in an up to 75% drop in advertising revenue for major players like Google and Facebook (Gill et al., 2013), and this obviously affects the advertising ecosystem as well as business trends. However, as mentioned in Section 10.4.3, SNSs may start to officially charge for their services if users do not wish to be tracked. The so-called “free communication services” is deceptive in reality because users are trading their information for the services offered by SNSs. Future work may discern whether SNSs are willing to do this, or whether SNS popularity will endure once the service is no longer free.

10.4.5 ICT professionals and the media

Another factor to be discussed is the contribution of researchers across different fields within the scope of ICTs (e.g., technical, social, and philosophical) and the role of the media. Technical accounts of privacy report on the technical flaws that may impinge on users’ right to privacy (e.g., Krishnamurthy and Wills (2010a); Humphreys, Gill, and Krishnamurthy (2010); Mayer (2011)) whereas philosophical accounts of privacy (e.g., Moor (1997b); Floridi (2005); Nissenbaum (2010); Tavani and Moor (2001)) provide insights into an issue from a moral and ethical perspective. Meanwhile, the findings from the social aspect of research that involved users’ perspectives are also crucial to report on users’ awareness and perception about how they manage their privacy online (Pempek et al., 2009; McDonald & Cranor, 2010b; Al-Saggaf, 2011). Therefore, a collaboration among researchers from these areas would be beneficial in pointing out the technical flaws while also explaining how they may impinge on users’ privacy, at the same time reporting on users’ awareness and concern over those flaws.

However, Krishnamurthy (2010) points out that the contributions from researchers on the topic of online tracking are usually ignored by the public; hence, tracking is still happening and keeps growing. He also adds that a Wall Street Journal article (Steel & Vascellaro, 2010) caused public breakouts and eventually triggered the response from Facebook’s founder about various privacy issues. Hence, both the researchers in ICTs and the media are seen to be effective in bringing the publicity’s attention to any issues brought by the technology and causing a change or response from the involved organisations such as Facebook.

10.4.6 Policy

Last but not least, another aspect to be considered is the policy design and enforcement regarding the collection and the retention of users’ information. Commissions in the United States, privacy commissioners in Canada, and the European Union, for example, have been concerned
about privacy online and there is also a proposed new law that requires first party sites to receive consent before placing a cookie on a user’s computer in the European Union (Krishnamurthy, 2010). This could be challenging because different cultures may hold different views regarding the importance and value of privacy. Although Cambodian participants expressed concern about privacy loss from online tracking, there is no comprehensive online data protection or privacy law in Cambodia (Anonymous, 2012); hence, there is no restriction on data collected online.

10.4.7 So?

Based on the rationale above, no single measure appears to work. For example, informed consent alone is not adequate, because some users may prefer not to have any recorded information about their online browsing. As in the three case scenarios discussed earlier, Soriya can choose to benefit from being a loyal customer of a supermarket while having recorded information about her and her shopping. Otherwise, she can choose to purchase goods and pay by cash and avoid having any information recorded. Should we move back to the olden days when paying by cash was more common? Electronic payment transactions are of course more convenient for both customers and the supermarkets. However, this could be a user’s option if they do not wish to be tracked.

However, this increasing flow and collection of information provides benefits to both online users and businesses. For example, business organisations like SNSs benefit from online advertising; hence, they are able to provide free contents and services which enhance online users’ experiences (FTC Staff, 2010). In addition to SNSs’ transparent policies, SNSs should also communicate to users the value of this type of information collection. Therefore, the approach to address user privacy should be flexible enough so that it allows companies to innovate their products and services without seriously limiting the benefits (Weckert, 2001; FTC Staff, 2010).

For the changes to occur, it requires an ongoing and long-lasting collaboration with and involvement in different interest groups such as the first and third party sites, online users, ICT researchers across different fields (e.g., technical, social and philosophical) who may discover the issues overlooked by the responsible organisations, the media that helps to spread the word, and the government who holds the power to impose the laws and rules.
Chapter 11

Conclusion

This chapter summarises the findings of this thesis while also commenting on possible future research. The chapter is organised as follows. Section 11.1 provides a summary of this thesis by restating the research questions and outlining the findings, and Section 11.2 describes the scope and limitations of the research. Lastly, Section 11.3 provides comments on future research that may arise from this research.

11.1 The research questions and findings

This research consisted of three main stages that aimed to address three main research questions. The first stage involved Wireshark experiments where researcher’s and participants’ online activities were recorded and analysed in order to investigate the ramifications of information sharing in HTTP headers. The second stage of the research was characterised by an ethnographic method and employed interviewing techniques to gauge participants’ perception and awareness of privacy and online tracking. Finally, the third stage of this thesis employed CI as a framework to analyse the practice of tracking online users by SNSs from a moral perspective. Research questions and findings in the thesis are outlined below.

1. What is the nature of information sharing in HTTP headers?
   (a) What type of information is being shared?
      First party sites visited by the researcher (in the exploratory study) and the participants in the study shared identifiable (e.g., name, and SNS unique identifier) and non-identifiable information (e.g., clickstream or browsing activities) to third party sites. This type of information is not intentionally shared to those third party sites by users.
(b) With whom is this information being shared?
Users’ unintentionally shared information traverses from first party or directly visited sites to third party sites. Some third party sites are advertisers or data aggregators (e.g., DoubleClick [Google’s advertising franchise], Scorecard research, and Yieldmanager [Yahoo’s advertising franchise]). Other third party sites are popular SNSs such as Google, Facebook and Twitter.

(c) What are the implications of this sharing?
It is very common for first party sites to use third party advertisers to display contents on their web pages. While retrieving advertising content from the advertising servers, first party sites were seen to share users’ clickstreams or names (e.g., LinkedIn case) to third party sites. Currently, it is also popular for first party sites to use SNS widgets (e.g., Facebook’s Like button, Twitter’s Tweet button, and Google’s +1 button) on their pages. Google, Facebook and Twitter are able to acquire information about users’ currently viewed pages among sites embedding SNS widgets. This study also found that users’ browser toolbars also connect to third party sites. While users’ browsers connect to the same third party sites, initiated either by first party sites or toolbars, those third party sites are able to track users’ movements across different first party sites via the use of the same cookies.

2. What are SNS users’ views on online tracking and privacy?

(a) What does privacy mean from SNS users’ perspectives?
From the Cambodian participants’ points of view, the meaning of privacy revolves around the notion of an individual’s right and control of others’ limited access to his/her information. The Relationship is seen to be the factor which determines whether a certain type of information is private. For example, family matters which are considered private in Cambodian culture - as well as among the majority of the participants - and can be disclosed or shared with close friends but not with the public or general SNS friends.

(b) To what extent are they aware of online tracking in general as well as online tracking by SNSs?
Almost all of the participants in the study were not aware of the tracking by the third party sites. Their perception of online tracking lies in the supposition that trackers are other Internet users who glean their information. All participants were well aware of the privacy settings in SNSs that can be used to limit the audience for their posts. They perceived that their online activities were private as long as they did not tell others what they do, but their activities on SNSs were not private as
other SNS users were able to see their posts. This finding suggests that the level of awareness of online tracking among Cambodian participants is relatively very low compared to participants from other countries as discussed in the literature.

(c) What are SNS users’ reactions to the practice of sharing and tracking information online?
The experience with the Wireshark experiments was new to them, and they were shocked at the outcome which indicated that their online activities were not private and could be recorded or collected by third party sites, including SNSs. They felt sad and angry and expressed concern about privacy loss, as this is happening without their consent or knowledge. The majority of subjects did not read privacy policies claiming they would not have any choice but to agree with the terms and conditions in order to create an SNS account, while a minority of the participants who read the policies reported that the policies were too long and too difficult for them to understand.

3. What are the ethical implications raised by the act of collecting and tracking users’ information by SNSs?
The CI framework was used to analyse the current practice of sharing and tracking of users’ information by SNSs. The framework rejected the practice for it violates the user’s right to privacy. SNSs did not gain users’ informed consent in moving or collecting their information across different contexts (SNSs and advertising contexts, and SNSs and non-SNS contexts). Users did not have the autonomy to decide when to go public, or when to disclose their unintentionally shared information to SNSs. SNS widgets were regarded as hidden cameras that are used to monitor users’ activities within and outside SNSs.

11.2 Limitations of research

The findings of the first stages of this research were based on the researcher’s and 20 participants’ online browsing activities. The researcher only investigated information sharing in HTTP headers. Online users’ information can be available to an unintended audience in other ways such as via flash cookies, or other robotic applications that crawl Web site contents. The study did not observe the browsing via other devices such as smartphones or tablets, but purely on personal computers. Online tracking can be sophisticated and therefore the findings of this study cannot be used to reflect online tracking in general, nor can it be used to reflect the whole population of online users.

The second stage of research was influenced by the ethnographic method. Twenty Cambodian
participants were interviewed about their perception and awareness of privacy and online tracking. Again, these findings are not generalisable to the whole population of Cambodian online users, nor to online users anywhere else. However, this revealed that in Cambodia where users have access to technologies such as smartphones, tablets, and computers, they may not be aware of the impacts these technologies bring to their lives. Yet, they did care about their privacy and they were concerned about the fact that their online activities could be tracked without their permission or informed consent.

Last but not least, the CI framework helped to shed light on the expectation of privacy in the context of SNSs and online tracking. This framework provided guidance through the different stages of analysis; yet it is not a single method or tool to analyse an issue. Tracking by SNSs could also be analysed from a legal perspective, but it did not fall within the scope of this research. In addition, there are other philosophical frameworks that can be used to analyse the impacts of technology on society. As in Soriya’s scenario in Chapter 1, the impacts of online tracking can also be analysed by using classical theories of utilitarianism and deontology, as well as other privacy frameworks like the ontological theory of informational privacy introduced by Floridi and RALC theory introduced by Moore and Tavani. Employing different frameworks as tools for analysis may yield different expectations and outcomes.

Despite the limitations, this research made a number of contributions. Firstly, it examined the information sharing and gathering in the HTTP headers resulting from the researcher’s and research participants’ online browsing activities. Secondly, it also studied the impact of this practice from the perspective of Cambodian online users, then from a moral perspective. The outcomes from both perspective are consistent: tracking by SNSs appears to be wrong. Online tracking has been around since the 1990s, but the tracking by SNSs is relatively new and has not yet been discussed much in the existing literature. Unlike the existing literature that tested a large number of cases, information sharing examined in this research was based on smaller number of cases, but included real life online users’ browsing activities. The existing literature has not yet studied Cambodian participants’ opinions on privacy and online tracking, nor did it studied the impact of SNS tracking from the CI perspective. This research also raised awareness among participants in Cambodia regarding online tracking - during the workshop where they were introduced to basic HTTP and during the interviews where they were revealed the experiment outcomes. It also paves the ways for future research that may derive from this research findings (more details in Section 11.3).
11.3 Future research

The main findings of this research suggested that there is a gap between the practice of tracking online users by SNSs and participants’ perceptions of such a practice. Future work may also determine whether the gap can be narrowed; for example, by increasing users’ awareness of the practice. As identified in Section 11.2, this research did not investigate information sharing in HTTP headers on devices other than personal computers. Smart phones and tablets are becoming popular and affordable; hence, there could be an investigation of information sharing on those devices, from the angle of real life users’ online browsing.

This research did not further explore why the majority of research participants were unaware of online tracking despite the fact that they had been exposed to smart gadgets (e.g., computers, mobile phones, and tablets) and the Internet for a number of years. Participants’ poor understanding of privacy policies or terms of services could also have contributed to their lack of understanding about the collection and use of their personal data. Future research may also investigate more deeply about the availability (or lack thereof) of comprehensible terms of services in regards to consent for non-English speaking users. This concern rises to significance from legal, ethical and social perspectives. In addition, future work may also examine whether the country’s economy, or online users’ academic or working background, gender and age, has an effect on users’ understanding of online tracking or other issues online.

The same study can be also repeated in a Western culture like Australian. This future study will allow a comparison between Western and non-Western user attitudes and the third party sites found during their browsing. In addition, as mentioned earlier in Section 11.2, this research employed only the CI framework. Different frameworks or theories may give different outcomes on the same practice. Future work may establish whether the same practice is acceptable from different theoretical viewpoints.

Future research may also investigate the propagation of unintentionally shared information, to determine the relationships or agreements between those third party sites. In addition, knowing that browser toolbars also instigate information sharing to third party sites could inspire technical work to contribute to the study of possible leakages via other browser adds-on while also discerning which strategies will reduce these leakages. Furthermore, this research did not focus on how users’ information will be used once it is collected. Future research may also investigate the practice of online tracking from the perspectives of other stakeholders like SNSs, advertisers, and data aggregators, in order to find out if this collection is harmful to users’ information, or identities.

As suggested in Section 10.4.3, and Section 10.4.4, the term “free communication services” ap-
pears to be deceptive, as users are trading their intentionally and unintentionally shared information for the service provided by SNSs. Future work may examine whether SNSs would still be popular if they began to charge for their service, giving users the option of having a private profile as well as private browsing. The researcher hopes that future work will help to reduce the negative impacts brought by this technology while at the same time maximising its potential benefits.
References


150


ference on social computing (socialcom) (p. 295-302). Alexandria, VA: IEEE.


disclosure, impressions, and interpersonal evaluation: getting to know one another a bit at a time. *Human Communication Research, 28*(3), 317 - 348.


Appendix A

Human Research Ethics Committee Approval
14 September 2012

Ms Rath Kanha Sar
School of Computing and Mathematics
WAGGA CAMPUS

Dear Ms Sar,

Your research proposal entitled “Privacy from the perspective of Cambodian social network site users: technical, cultural and philosophical perspectives” has been reviewed by Charles Sturt University’s (CSU) Human Research Ethics Committee (HREC).

The CSU HREC operates in accordance with the National Health and Medical Research Council’s National Statement on Ethical Conduct in Research Involving Humans.

I am pleased to advise that the project meets the requirements of the National Statement; and ethical approval for this research is granted for a twelve month period from 14/9/12.

The protocol number issued with respect to this project is 2012/144. Please be sure to quote this number when responding to any request made by the Committee.

Please note the following conditions of approval:

- all Consent Forms and Information Sheets are to be printed on Charles Sturt University letterhead. Students should liaise with their Supervisor to arrange to have these documents printed;
- you must notify the Committee immediately in writing should your research differ in any way from that proposed. Forms are available at http://www.csu.edu.au/_data/assets/word_doc/0010/176833/ehcr_annrep.doc;
- you must notify the Committee immediately if any serious and or unexpected adverse events or outcomes occur associated with your research, that might affect the participants and therefore ethical acceptability of the project. An Adverse Incident form is available from the website; as above;
- amendments to the research design must be reviewed and approved by the Human Research Ethics Committee before commencement. Forms are available at the website above;

Version 2

www.csu.edu.au

ACNC registered charity number: 000000995

PRA

ABN: 43 878 798 551

157
• if an extension of the approval period is required, a request must be submitted to the Human Research Ethics Committee. Forms are available at the website above;
• you are required to complete a Progress Report form, which can be downloaded as above, by 14/9/13 if your research has not been completed by that date;
• you are required to submit a final report, the form is available from the website above.

YOU ARE REMINDED THAT AN APPROVAL LETTER FROM THE CSU HREC CONSTITUTES ETHICAL APPROVAL ONLY.

If your research involves the use of radiation, biological materials, chemicals or animals a separate approval is required from the appropriate University Committee.

The Committee wishes you well in your research and please do not hesitate to contact the Executive Officer on telephone (02) 6338 4628 or email ethics@csu.edu.au if you have any enquiries.

Yours sincerely

[Signature]

Julie Hicks
Executive Officer
Human Research Ethics Committee
Direct Telephone: (02) 6338 4628
Email: ethics@csu.edu.au
Cc: Dr Yedan Al-Sugfah
Appendix B

Information sheet
Participant information sheet

Privacy from the perspective of Cambodian social network site users: technical, cultural and philosophical perspectives.

My name is Rath Kanha (Lis) Sar and I am a PhD candidate at the School of Computing and Mathematics, Charles Sturt University, Australia.

I would like to invite you to take part in my research which aims to study people’s experience in terms of privacy management and awareness in using social network sites (SNS). This project is supervised by Dr. Yeslam Al-Saggaf (yalsaggaf@csu.edu.au).

Social Network Sites (SNS) such as Facebook, Twitter, and LinkedIn, Google Plus, MySpace and Friendster provide new means of communication and also bring exciting opportunities, but at the same time raise concerns about privacy, particularly the risk of personal information leakages and the secondary use of those pieces of information.

The users’ information can be leaked or become publicly available in a number of ways including through the privacy setting in SNS (e.g. the user does not know how to manage their SNS privacy setting properly, thus the profile information becomes publicly available by default), the use of the tracking technologies such as HTTP cookies (HTTP cookie is a small piece of text file stored by the browser at the user’s machine, for example, to remember the content in a user’s shopping cart), and the use of the Application Programming Interface (API) among SNS sites (e.g. the sharing of the user’s information to the third party application like Farmville in Facebook).

I am looking for 20 participants who meet the criteria described in (3) to participate in my study by attending a short training session, and who are comfortable enough to talk about their online experience in a face-to-face interview.

It is hoped that the study will help us understand better what privacy means to SNS users, how important it is to them, how the users manage their privacy online and what their reactions are about what is happening in terms of information leakages.

1. What is the purpose of the study?
   The study aims to understand the meaning and importance of privacy from the perspective of Cambodian SNS users. This study will first let the participants witness how SNS and non-SNS track their online movement in the HTTP headers (or the messages exchanged by the sites while communicating with each other) by using Wireshark as a tool. Next, the participants will be shown the results and the implication of the tracking and they will be interviewed by the chief investigator about their feeling and opinion regarding the meaning and importance of their privacy online.

2. What does this study involve?
   It involves two stages: Wireshark experiments and a face-to-face interview – both are conducted in English language.

   The Wireshark experiment involves each participant performing online browsing activities (according to the experiments procedures given during the introductory workshop) while having these activities recorded by Wireshark. The participants will be required to attend an introductory workshop about Wireshark and HTTP before the experiment begins.

   Next, each participant will be interviewed face-to-face by the chief investigator. The Wireshark records will be analysed by the chief investigator, and the results will be given to each participant during interview.

3. Who can participate in this study?
   To participate you must:
   - Be Cambodian,
   - Be an academic staff or enrolled as a student at the Institute of Foreign Language (IFL), English department,
   - Be from generation X (born 1965-1980) or Y (born 1981-1994),
   - Have at least 1 Social Network Site (SNS) account or profile (e.g. Facebook, Twitter, LinkedIn, MySpace, or Friendster …)
- Use internet on a regular basis (e.g. Checking emails, browsing on the webs…) 

4. **What are the benefits of participating in this study?**

You will be given an introductory workshop on what HTTP is and how to use Wireshark to capture your network traffic between your browser and the requested sites. In addition, your opinion regarding privacy will be heard!

5. **Are there any possible risks from participating in this study?**

There is a possibility that your password can be revealed in Wireshark if it is not properly encrypted (very unlikely from well-known sites like Facebook, or Google), so you are advised to change your password before and immediately after the experiment.

If at some stage during the interview, you become uncomfortable or distressed, you are free to decline to answer any question or to end the interview, and all the previous conversation and your Wireshark dump files will be destroyed.

6. **What type of data/information collected about me?**

Wireshark dump files during the experiment and your opinion in the interview are crucial for the research. However, any identifiable information about you will be replaced by a pseudonym or a nickname. All the information and the conversation will be confidential and safeguarded, and will not be shared with any third party organisation.

7. **How do I participate in this study?**

To find out more information or to register your interest in taking part in the study, please feel free to contact the chief investigator at rsar@csu.edu.au.

Thank you so much for taking time to consider and participate in this study ☝️

Charles Sturt University’s Human Research Ethics Committee has approved this study.

---

Executive Officer
Human Research Ethics Committee
Office Academic Governance
Charles Sturt University
Bathurst NSW 2795

Phone: (02) 6338 4628
Email: ethics@csu.edu.au
Appendix C

Consent form
Consent form

Project title: Privacy from the perspective of Cambodian social network site users: technical, cultural and philosophical perspectives.

Chief investigator: Rath Kanha (Lis) Sar (rsar@csu.edu.au)

Supervisor: Dr. Yeslam Al-Saggaf (yalsaggaf@csu.edu.au)

Charles Sturt University,
Boorooma Street, Locked Bag 588,
School of computing and mathematics,
Wagga Wagga NSW 2678
Australia.

I _________________________ agree to participate in the study titled above.

I understand that my participation is voluntary; that I can choose not to participate in part or all of the project and that I can withdraw at any stage of the project.

I have read and understood the purpose of the project and I understand that the study will be conducted in English language.

I understand the project involves two stages: the Wireshark experiment and the interview, and that by signing below I agree to participate in both.

I understand that Wireshark files and any conversation in the interview will be used as part of the project.

I understand that any information provided is confidential, and that no information that could lead to the identification of any individual will be disclosed in any report on the project, or to any other party.

I understand that the reports based on the Wireshark experiments and interview will be kept in a secure storage and accessible to the chief investigator and her principle supervisor only. I also understand that the reports held by the University will be destroyed after five years from the completion of the study.

I understand that login details such as password can possibly be revealed in the Wireshark file and I have been informed to change my password before and immediately after the experimental study.

Charles Sturt University’s Human Research Ethics Committee has approved this study.
I understand that if I have any complaints or concerns about this research I can contact:

Executive Officer
Human Research Ethics Committee
Office of Academic Governance
Charles Sturt University
Panorama Avenue
Bathurst NSW 2795

Phone: (02) 6338 4628
Email: ethics@csu.edu.au

Signature: ________________________________
Date: __________________________________
Interviewer’s Name: Rath Kanha (Lis) Sar
Appendix D

Interview Questions
Individual Interview Questions & Responses

Date: _____________________________
Location: _____________________________
Interviewee: _____________________________

1. What are the device(s) and the browser(s) you usually use to go online?

2. What do you usually do online?

3. What are the SNS you are currently with, and SNS you used in the past (if applicable)?

4. How often do you visit those sites?

5. For how long or since when have you been using SNS?

6. What motivates you to create an SNS account?

7. What types of information are you sharing on SNS?

8. What motivates you to share those types of information?

9. What do you think of the privacy policies in the SNS?

10. How do you manage SNS privacy setting in terms of ‘who’ can view your profile or other pieces of information?
11. What does privacy mean to you?

12. What types of information or activities do you consider private or personal?

13. How important is your privacy?

14. What is your view regarding your online activities? How private do you think your online activities and information shared on SNS are?

15. What is your view on the fact that most of your online activities are being tracked/recorded?

16. How do you feel about the fact that someone can create a digital profile of you, which includes your SNS profiles and your online activities?

17. What are the technologies or techniques or tools do you use to lessen the chance of your information tracking?
Appendix E

List of participants
<table>
<thead>
<tr>
<th>N</th>
<th>Names</th>
<th>Gender</th>
<th>N</th>
<th>Names</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dara</td>
<td>Male</td>
<td>11</td>
<td>Tola</td>
<td>Male</td>
</tr>
<tr>
<td>2</td>
<td>Kosal</td>
<td>Male</td>
<td>12</td>
<td>Phara</td>
<td>Male</td>
</tr>
<tr>
<td>3</td>
<td>Rany</td>
<td>Female</td>
<td>13</td>
<td>Bopha</td>
<td>Female</td>
</tr>
<tr>
<td>4</td>
<td>Phary</td>
<td>Female</td>
<td>14</td>
<td>Seila</td>
<td>Male</td>
</tr>
<tr>
<td>5</td>
<td>Seyha</td>
<td>Male</td>
<td>15</td>
<td>Phanna</td>
<td>Male</td>
</tr>
<tr>
<td>6</td>
<td>Tevy</td>
<td>Female</td>
<td>16</td>
<td>Sitha</td>
<td>Female</td>
</tr>
<tr>
<td>7</td>
<td>Moly</td>
<td>Female</td>
<td>17</td>
<td>Rithy</td>
<td>Male</td>
</tr>
<tr>
<td>8</td>
<td>Nary</td>
<td>Female</td>
<td>18</td>
<td>Mony</td>
<td>Male</td>
</tr>
<tr>
<td>9</td>
<td>Makara</td>
<td>Male</td>
<td>19</td>
<td>Vibol</td>
<td>Male</td>
</tr>
<tr>
<td>10</td>
<td>Nara</td>
<td>Male</td>
<td>20</td>
<td>Udom</td>
<td>Male</td>
</tr>
</tbody>
</table>

Table E.1: List of participants