32144 Technology Research Preparation

Week 5 Tutorial 4 Overview	
Topics Covered:	Synthesizing the Literature Themes & Synthesis Framework Synthesis Matrix
Points:	6
Date Due:	Week 5 – In Class
Submission	In-class contribution to discussions during Week 5 tutorial session

Week 5 Tutorial Preparation – Preparation of Assignment 2 Introduction.

TRP - Assessment Task 2	
Criteria	
INTRODUCTION Presents the problem, sets up the field, and states the student's point of view; Clearly stated and well-written aims, objectives and significance of the knowledge contribution of the review; Orients the reader to the report by providing an overview.	**20 to >17.0 pts** **MASTERY** Superior statement of the research problem. States aim of research and communicates importance of topic. States main findings. Brief but complete overview of report organisation.

On the following page, write a draft Introduction for your literature review assignment based on your analysis of the rubric. We covered this in Week 4 lecture.

Your tutor will arrange either a presentation and discussion tor your group or to the tutorial class.

These bullet points were derived from the Rubric of Assignment 2 (opposite).

Your single page draft should include the following information:

- State the Research Problem
- State aim of research
- Importance of the research
- State Main Findings – what is the knowledge gap that you have discovered?
- Brief but complete overview of report organization.

These bullet points ensure that you meet the requirements of the rubric.

Write Your Draft Introduction to Assignment 2 Literature Review.

Computer networking is the crux of the majority of relied upon technologies of the modern age, notably the internet. As of August 2023 [IBIS report] ISPs in Australia was a 6.5 billion dollar industry, alluding to its high usage in modern society. Consequently, databases interacting on such networks become susceptible to attacks and confidential and proprietary information is put in potential jeopardy of being disclosed to undesired parties.

The 2014 Sony a data breach conducted by the North Korean intelligence group 'the Lazarus group' exposed roughly 30 million files worth of salacious content, private emails between executives and celebrities [Mills et al.] to the public, resulting in an estimated loss of billions of dollars in revenue.

However the most noteworthy example is the 2013 and 2014 Yahoo! data breaches occurring due to inadequate hashing procedures and susceptibility to phishing schemes orchestrated by the Russian FSB [Daswani et al.]. These resulted in the public disclosure of 500 million user's credentials and affected all 3 billion users of Yahoo!'s services.

In response, mathematicians and computer scientists alike have developed a corpus of literature discussing the optimization of mathematical methods in the context of such data breaches, with breakthrough results, notably in cryptography through number theory and group theory, as well as anomaly detection through probability theory and AI.

Despite considerable progress, there exist knowledge gaps in the current state of literature relating to unknown or unrefined solutions that have practical shortcomings. With the resurgence of AI and the repertoire of modern tools recently introduced, anomaly detection has the potential to be greatly improved, especially when contrasted to Bayesian-based techniques. On the converse, developments in computing are constantly placing cryptography in jeopardy of being computationally feasible to break (such as the MD5 hashing system); particularly in the light of quantum computing, demanding alternatives or refinements to uphold the breakthroughs of cryptography achieved by research.

Few articles provide relation of vulnerabilities with previous case studies of data breaches, which plays a crucial role in demonstrating the utility of mathematical research outcomes (which can appear abstract on first introduction) and their capability of being employed to protect the data relied upon by organizations and the general public.

The scope of this literature report concentrates on data breach prevention from a mathematical perspective; hence articles relating to data breach history or have a mathematical nature with applications in cybersecurity will form the basis upon all points of discussion.

This report will synthesise an initial critical analysis concentrated on two outstanding articles related to the topic, drawing constant comparisons with a metric to evaluate the level of topic relevance, reliability, accuracy, potential for bias, timeliness and completeness exhibited by each article.

Additionally, a cohesive literature review will depict the broader state of the focus' corpus, discussing how recent research has contributed to the optimization of mathematical methods for data breach prevention, and highlighting critical knowledge gaps whose satisfaction will motivate the advancement of cybersecurity research.

A conclusion will summarise the key ideas representing the literary corpus and allude to the direction in which the progression of data breach prevention research may take considering the observations in literature.

Week 5 Tutorial Preparation – Reading and Analyzing Papers

1. **CONTINUE IDENTIFYING AND READING PAPERS** – You should have identified at least seven relevant research articles from the academic and/or industry databases by the end of this week and be finalising the main sources you will use in your literature review. You should use academic or industry databases (such as ProQuest Central, Google Scholar, *Scopus* (Elsevier) *Web* of Science, and Discovery (EBSCO)) or industry report databases (Gartner.com, IBIS Industry World Reports), which are available via the UTS Library website (Refer to Week 3 Lecture).

2. **FINALISE YOUR MOST RELEVANT PAPERS FOR YOUR REVIEW** – This week you should be identifying the **most relevant papers you have sourced to date**. From the sources you have collected, begin to shortlist conference papers, journal articles and industry reports. You do not need to have this task completed before your tutorial in Week 5, however you should **have identified 4-5 papers** that you can use to develop ideas for your synthesis framework (see below).

   THINK ABOUT AND REFLECT ON MAIN THEMES – Using your research question and your most relevant papers, identify a few of the **major themes** that you can use to critical evaluate and analyse your sources. You have already created search criteria for selecting the papers, which is a starting point for your major themes. Now you need to go deeper into those sources to identify your major themes, which should be identified relative to the concepts that define your topic, and can be based on e.g., Linkages and connections, Trends and similarities, Contradictions and contrasts, and Causes, factors, variables, etc.

3. **READ THE PAPER BY DENNEY AND TEWKSBURY (2013)**: You will recall from Week 5 that there are different purposes and approaches to writing a literature review. Denney & Tewksbury (2013) highlight the three primary functions of a literature review: **integrative**, **theoretical**, or **methodological**.

   > "Reviews may be **integrative** (summarizing past research based on overall conclusions of the past research), **theoretical** (identifying and critiquing the ability of different theories to explain a phenomenon), or **methodological** (highlighting different methodological approaches used in past research and the contributions of each type of research) in focus."[1]

[1] Denney, A. S., & Tewksbury, R. (2013). How to write a literature review. Journal of criminal justice education, 24(2), 218-234.

In-Class Tutorial Activities

1. EXPLORING YOUR MAJOR THEMES

ACTIVITY 1: Identify and Discuss Your Major Themes – In breakout groups, identify themes that you can use to critically evaluate and analyse your sources. As stated above, you should identify themes relative to the concepts that define your topic, and can be based on e.g.,

- Linkages and connections,
- Trends and similarities,
- Contradictions and contrasts,
- Causes, factors, variables, etc.

Topic: Databreach prevention from a mathematical perspective

Major Theme 1:

Anomaly detection

- Discussing knowledge gaps in anomaly detection regarding solutions that can be significanly refined using modern techniques (such as applications of modern AI)

Major Theme 2:

Cryptography

Discussing knowledge gap regarding unresolved issues in cryptography

Major Theme 3:

Anecdotal demonstration of relevance

- Relating the previous problems and solutions towards real case studies of data breaches

- Will promote the importance of the topic of research and profondity of solutions, linking the theoretical results to practical benefits

NB: Your major themes should frame your research problem and provide supporting arguments for your research question, statement and purpose, and ultimately define the objectives of your research proposal (Task 3). Figure 1, shows this funnelling effect that your major themes should support.

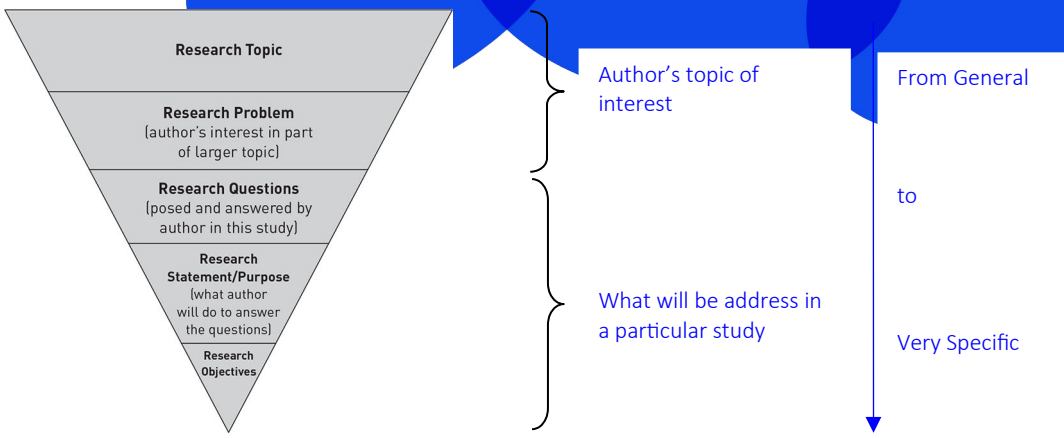


Figure 1. Research Topic, Research Problem, Research Question, Research Statement/ Purpose, Research Objectives. The research problem is the author's topic of interest within a larger area of interest (research topic), and the research question and the research statement reflect what will be addressed in a particular study. In other words, authors move from the general to very specific as they set up their paper. The Introduction section of the papers you are reviewing will contain all of these, and the good ones should achieve this as an inverted triangle or cone.

2. DEVELOPING YOUR THEME & SYNTHESIS FRAMEWORK

ACTIVITY 2: Identify and Discuss the Relationships between Your Major Themes – Using the major themes that you have identified from the literature identify any relationships between them and create a relationship diagram to map any associations between them.

YOU MIGHT USE RELATIONSHIP DIAGRAMS SUCH AS:

a. Cause and effect diagram
b. Venn diagram (see example in figure 3 below)
c. Hierarchical diagram
d. Process / cycle diagram … etc

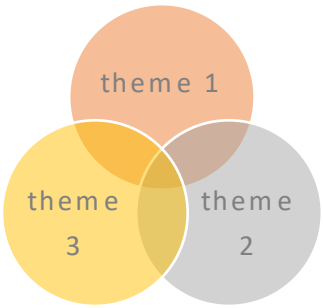
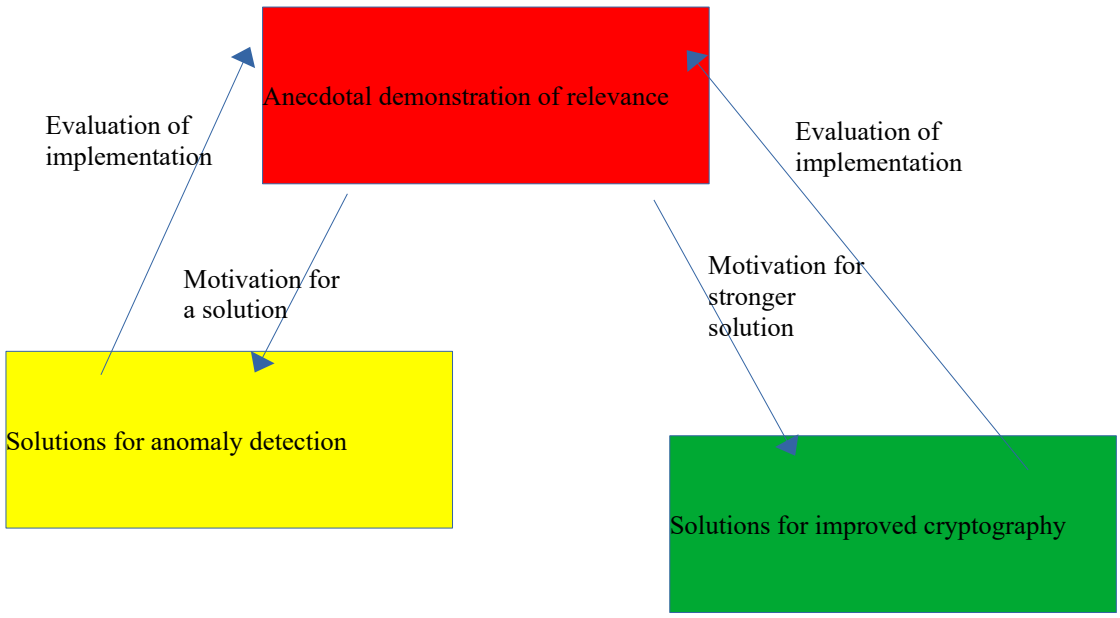


Figure 3. Venn Diagram

Tree diagram relating topics



The topics are related by their common motivator; the case studies (anecdotes).

Once such a solution in either anomaly detection or cryptography is finally employed, its practical result can be evaluated and discussed in a future literature review

3. COMPLETING THE SYNTHESIS MATRIX –

ACTIVITY 3: Register and then Discuss a Source Article – Complete the synthesis matrix for one or even two articles during the tutorial (depending on time). Before then sharing the results in a feedback discussion.

NB: Please use APA Referencing (Authors, Date, Title, etc.) for your selected conference paper, journal article or industry report.

NB: You can use the matrix below or use the excel spreadsheet that is available on Canvas.

In-Class Tutorial Activities – USE THIS SYNTHESIS MATRIX TO SYNTHESISE THE ARTICLES

SOURCE (Author/s)	SOURCE A (E.G., DENNEY & TEWKSBURY)	SOURCE B Machine learning for email spam filtering: review, approaches and open research problems	SOURCE C A Study of the MD5 Attacks: Insights and Improvements	SOURCE D	SOURCE …
YEAR	2013	2019	2006 (On the older side, however I believe this is still a noteworthy, relevant article to discuss)		
MAIN CONCEPT/ THEME	Here I list the **main concept or the major theme of the paper** that is specifically related to my topic and my literature review	Anomaly detection	Cryptography		
OVERVIEW	Here I **summarise the article in a sentence or two**. It can be based on the abstract but should be more specifically related to the article's relevance to your research topic.	Primarily the introduction of machine learning concepts in anomaly detection to create stonger spam filters and comparison to Bayesian methods of spam filtering	The exposure of the weaknesses of the MD5 hashing algorithm by constructing an algorithm that finds collisions.		
METHODS	Here I summarise the **research methods used** (e.g., interviews, surveys, experiments, literature review).	Experiments were used to verify superiority over Naiive bayesian spam filters. Literature review of previous spam filtering ideas is also employed.	Literature review of Wang's algorithm for collision detections is the primary method of research, however surveying amongst authors for improvements to the algorithm for time complexity is another method.		
RESEARCH DESIGN	Here I describe the **approach to the research** (qualitative or quantitative)	Quantitative; uses statistical methods for comparison of newly proposed technology in contrast with bayesian methods	Qualitative; Method using for finding collisions was based on observations of each author while coding (technically surveying and hence according to Wakefield's definition, qualitative)		
Anomaly detection Based on Major Themes, Connections, Causes, Factors, etc. to your Research Topic/ Question	Here I make notes about **a specific theme in the article that is relevant to my research topic**/ question. I usually try to insert a quote or an image that relates to this theme. + Pg. No. I may also note the extent that I agree or disagree.	Provides a pioneering solution of applying machine learning techiques to anomaly detection	Not applicable; this article does not deal with anomaly detection; see the relation diagram above.		
Cryptography Based on Major Themes, Connections, Causes, Factors, etc. to your Research Topic/ Question	Here I make notes about a **second theme in the article that is relevant to my research topic/ question**. I usually try to insert a quote or an image that relates to this theme. + Pg. No. I may also note the extent that I agree or disagree.	Not applicable; this article does not deal with cryptography; see the relation diagram above.	"The methods used by the Chinese team require an expected 2 37 MD5 computations to find the first block pair of the colliding		

			messages, and an expected 2 30 MD5 computations to find the second block pair. Klima [8] improved the attack so that an expected 2 33 and 2 24 MD5 computations are needed to find the first and second, respectively, message block pairs, although Klima did not implement his improved attack for finding the second block pair. Our method improves the attack so that an expected 2 30 MD5 computations are required to find the first block pair, and we implement Klima's code for finding the second block pair" The article shows vulnerabilities in the commonly used cryptographic scheme MD5. It implicitly poses the cryptographic challenge to design stronger algorithms for practical use.		
Anecdotal demonstration of relevance Based on Major Themes, Connections, Causes, Factors, etc. to your Research Topic/ Question	Here I make notes about **a third theme in the article that is relevant to my research topic/ question**. I usually try to insert a quote or an image that relates to this theme. + Pg. No. I may also note the extent that I agree or disagree.	'Different spam filtering formulas have been employed by Gmail, Ou tlook.com and Yahoo Mail to deliver only the valid emails to their users and filter out	Lack of relation to this topic is mentioned the article to real data breaches, however the topic has extreme relevance to the 2014 Yahoo! Data breaches where		

		the illegitimate messages. Conversely, these filters also sometimes erroneously block authentic messages. It has been reported that about 20 percent of authorization based emails usually fail to get to the inbox of the expected recipient. ‘ The following citation mentions the shortcomings of large Email providers that employ Naiive bayesian filters, this could be further elaborated on to link with the 2014 Yahoo! Data breaches which were predominantly conducted through phishing schemes which such research aims to prevent	MD5 hashes were used for credential storing despite being insecure.		
KEY REFERENCES	Here I list the references or citations that link the article to other research articles that I have identified as being important to my research topic.	Bhowmick, A., & Hazarika, S. M. (2016). Machine learning for e-mail spam filtering: review, techniques and trends. *arXiv preprint arXiv:1606.01042*. “Due to their resourcefulness, they are evolving as a major tool in the machine-learning researcher's set of tools. Nevertheless, neural networks	Wang, X., & Yu, H. (2005, May). How to break MD5 and other hash functions. In *Annual international conference on the theory and applications of cryptographic techniques* (pp. 19-35). Berlin, Heidelberg: Springer Berlin Heidelberg.		

		are not commonly used in the detection of spam email as one may possibly envisage. As an alternative, nearly all state-of-the-art spam filters use naïve Bayes classifiers.”			
CONCLUSIONS OR MAIN “TAKE-AWAYS”	Based on the paper’s conclusions, what are **the main take-aways** that are pertinent to my own Research Topic/ Question. I note the extent that I agree or disagree with the paper’s conclusions. I briefly comment why I agree/ disagree.	The paper concludes by stating thaty probability based spam filters can be extended upon by employing a combination of all machine learning techniques.	The paper concludes by summarising a method to find collisions for MD5 hashes described by Wang and their computational addition which shortens the running time of this algorithm.		
RESEARCH GAPS	Here I **identify the gaps** in the article. With regard to your research problem, what key issues has the article not addressed. You can identify them or find them in the section on limitations and areas for future research.	- Can be refined using new technologies in AI such as ResNN; knowledge gap stems from using more modern tools to develop the originally established ideas	- Opens possibility for new hashing functions to be promoted/developed as well as solutions to strengthen MD5 hashes against these methods (modifications to the MD5 standard)		
KIND OF LITERATURE REVIEW	Classify the **article’s literature review**. Here I note the kind of literature review the authors presented. It is either integrative, theoretical, or methodological	Integrative; extending on a chosen idea within the topic. The article explores the current solutions of Naiive bayesian methods and extends on these methods using modern technologies.	Methodological; demonstrating the state of the art technique relating to the topic. The article demonstrates the method of Wang for finding collisions for MD5 hashes. Although it does offer an extension to shorten the runningtime of Wang’s algorithm, it is not the focus of the article.		
PUBLICATION TYPE	Book/ Book Chapter/ Journal Article/ Conference Paper/ Government or Commercial Consultant Report/ Industry Journal/ Website Article	Journal Article	Journal Article		

References

Here I list the full **APA citation** (of the article, book, book chapter, etc. provided in the table above)

For example:

Denney, A. S., & Tewksbury, R. (2013). How to write a literature review. *Journal of criminal justice education, 24*(2), 218-234.

Bhowmick, A., & Hazarika, S. M. (2016). Machine learning for e-mail spam filtering: review, techniques and trends. *arXiv preprint arXiv:1606.01042*.

Black, John, et al. "A Study of the MD5 Attacks: Insights and Improvements." *Lecture Notes in Computer Science*, vol. 4047, 1 Jan. 2006, pp. 262–277, https://doi.org/10.1007/11799313_17. Accessed 23 July 2023.

Dada, Emmanuel Gbenga, et al. "Machine Learning for Email Spam Filtering: Review, Approaches and Open Research Problems." *Heliyon*, vol. 5, no. 6, June 2019, p. e01802, www.sciencedirect.com/science/article/pii/S2405844018353404, https://doi.org/10.1016/j.heliyon.2019.e01802.

Wang, Xiaoyun, and Hongbo Yu. "How to Break MD5 and Other Hash Functions." *Lecture Notes in Computer Science*, vol. 3494, 2005, pp. 19–35, https://doi.org/10.1007/11426639_2.