Dimitar Ivanov
Jacob Trock

1. Who are the *direct* and *indirect stakeholders* here? For each stakeholder, briefly explain the most important *values* they may have that are at stake in this discussion.

- Direct Stakeholders (and associated values)
  - Apple
    - Reputation, Potential Liability
      - Apple doesn't want to be a host to this kind of material, so they have an interest in screening it off their platforms to prevent associations or potential liability.
  - IMessage users
    - Privacy, security
      - Users wish to maintain their right to privacy, not have to worry about someone watching their every text and potentially abusing that privilege.
  - Criminals(illegitimate)
    - Crime
      - Criminals want to continue operating by distributing illegal materials without being caught.
- Indirect Stakeholders (and associated values)
  - Potential Victims/Victims
    - Safety and well being
      - Victims value safety and wish to know that illegal material of them isn't being circulated.
  - Host government/law enforcement
    - Effective use of resources
      - If false positives are rampant, then police resources could be wasted in their investigation.
    - Capturing criminals
      - The host government and law enforcement want to catch criminals, especially if they're actively commiting crimes and will continue to commit more crimes in the future.
  - Oppressive Regimes (Illegitimate)
    - Spying on citizens and silencing dissidents
      - If a system were made available to filter images based on the criteria of the host government, said government could abuse this by filtering for images such as anti-government memes.

Dimitar Ivanov
Jacob Trock

2. A crucial part of VSD analysis is technical inquiry. That involves a careful investigation of what technical features of a technology are possible, and how they might be implemented in a way that is consistent with important values. Based on what you have learned in this course so far (and on prior knowledge), do you agree with the criticism that Apple's proposed changes would break end-to-end encryption in its Messages app when that app is paired with an iCloud account? Why or why not?

- By strict definition, Apple will be breaking end-to-end encryption in the sense that it will be monitoring user's messages; however, they will only be checking for illicit content using private set intersection, a method in which no unobfuscated materials would be transmitted to Apple, rather just a hashes of the image(s) in question, which will not allow Apple to identify what the contents of non-offending images are. This could be done before the message encryption step, therefore it can be done without creating a backdoor in the encryption protocol. Due to the implementation, the likelihood of false positives is extremely low, however the possibility for false negatives is high.

3. Imagine that Apple decides to continue with its planned changes. Briefly describe some of the worst-case scenarios that could result. What bad thing(s) could happen? To which stakeholders? Which stakeholders' values or interests would be harmed?

- False positives(hash collisions) - If a bad hashing algorithm is used, then innocent users could be placed under investigation for something which isn't CSAM, leading to an invasion of their privacy and a waste of police resources. However, this is mitigated by using a good hashing algorithm and having a lower limit on the number of hash matches made before a report is generated.
- Governmental Abuse of technology - Oppressive regimes could attempt to force Apple to use the technology with their own set of hashes that pertain to locally prohibited materials, which could include things such as anti-government images, lgbtqia materials and materials of underground religious movements to name a few. This could subvert user's right to privacy and Apple's desire to maintain their reputation for privacy.
- Hackers using the check step as a backdoor - The existence of a system that can access the contents of messages before they're sent could be exploited to extract the contents of said messages without suspicion of the user. This could be used to blackmail users, spy on them (which violates their right to privacy) or generate false hash matches to incriminate the victim of the hacking attack (which also wastes police resources).

Dimitar Ivanov
Jacob Trock

4. It seems highly unlikely that a technical solution (e.g., a new cryptographic algorithm) can be developed that will resolve the debate over Apple's planned changes, in a way that satisfies all stakeholders. In the absence of such a solution, people will need to make tradeoffs between the values of different stakeholders.

If you were able to decide the outcome of the debate, what outcome would you choose, and why? In answering this, make sure to explain:

1. **which stakeholders and values your outcome favors,**
1. **which stakeholders and values are compromised; and**
1. **why these value tradeoffs are appropriate.**

Hint: Are any stakeholder values illegitimate in this context? Are rights that need to be respected or "lines" that cannot be crossed? And which of the remaining values are strongest?

We believe that if the system is executed properly, i.e. with strong hashing algorithms, a secure hash checking program, and without introducing encryption backdoors, the benefits of this change would outweigh the negatives. If done in this manner, the user's right to privacy would be virtually uninhibited, because none of their non-offending messages contents would be shared with Apple, and the protections of end-to-end encryption are preserved. The only people who would be compromised by action like this would be those that send CSAM images, whose stakes in the matter are illegitimate. This will help in the preservation of the right to safety of victims and potential victims. It will also protect their privacy by preventing the circulation of nonconsensual images of them. While Apple's and the governments' stakes are less important in the current context, this change will still accommodate their desires by, in Apple's case, preventing the illicit material from circulating on their platform, and in the government's case by assisting them in the pursuit of criminals in a way which efficiently uses police resources. The user's privacy will be slightly diminished by this, however because the flag will rarely be triggered, and only by cases of CSAM which are truly flagrant, non-criminal users will be virtually unaffected, making the theoretical reduction in privacy justified.

5. Thus far, our analysis has assumed a U.S. context. Would your proposal for how to resolve this debate (question 4) change if the debate was being held in a different country? Why or why not? What factors are relevant here?

While the system in America functions in alliance with the National Center for Missing and Exploited Children, which acts as a trusted source for hashes of CSAM, globally

there would have to be a different partner organization. If something of the sort doesn't exist already, there would need to be a global corpus of CSAM hashes that could be checked against content. This would prevent localized abuses of hash databases by authoritarian regimes that wish to supply hashes beyond just CSAM by forcing many nations to agree on the contents of that CSAM hash server and which of it will be flagged. This agreement would prevent suppression via this system and the problems caused by regional CSAM databases. There would have to be some sort of agreement system as well to prevent global abuses of the system, in the same sense as how interpol is sometimes abused by authoritarian nations.