Description of the experiment

- Two armies, each led by a general, are preparing to attack an enemy army.
- The armies are encamped each on its own hill.
- A valley separates the two hills, and the only way for the two generals to communicate is by sending messangers through the valley.
- The valley is occupied by the city's defenders and there's a chance that any given messenger sent through the valley will be captured.
- Two generals have agreed that they will attack, but they haven't agreed upon a time for attack before taking up their positions on their respective hills.

The two generals must attack the city at the same time to succeed. Having a single general attack is considered a disastrous failure. The problem is to come up with algorithms that the generals can use, including sending messages and processing received messages, that can allow them to correctly conclude that they will attack at the agreed time.

A possible sequence of messages

- General A: attack at dawn!
- General A: has the messenger been captured?
  - If so, I will attack alone; better if I don't attack
  - If not, general B will attack; better if I attack
- General B: General A is not aware of my receipt of the message; better if I send an "acknowledgement"
- General B: has the acknowledgement been captured?
  - If so, I will attack alone; better if I don't attack
  - If not, general B will attack; better if I attack

Proof of impossibility

Suppose there is any sequence of messages, some successfully delivered and some not, which suffice to meet the requirement of shared certainty for both generals to attack. In that case there must be some minimal subset of the successfully delivered messages that suffices (at least one message with the time/plan must be delivered). Consider the last such message that was successfully delivered in such a minimal sequence. If that last message had not been successfully delivered then the requirement wouldn't have been met and one general at least (presumably the receiver) would decide not to attack. From the viewpoint of the sender of that last message, however, the sequence of messages sent and delivered is exactly the same as it would have been had that message been delivered. Therefore the general sending that last message will still decide to attack. We've now constructed a circumstance where with the purported solution algorithms one general will attack and the other will not - contradicting the assumption that the algorithms were a solution to the problem.