7 Common knowledge and agreement

Common knowledge is a prerequisite for agreement. Namely, if a group comes to agree on something, then any person in the group knows that each member in the group agrees. Suppose that $agree(\varphi)$ is a formula that is true at the states in which the group has agreed on φ. Then, it is expected that

$$agree(\varphi) \Rightarrow E_G agree(\varphi)$$

is valid. The induction rule for common knowledge now implies that

$$agree(\varphi) \Rightarrow C_G agree(\varphi)$$

Hence, agreement implies common knowledge.

7.1 Coordinated attack

Two divisions of an army are camped on the opposite sides of a valley. In the valley is the enemy. If the two divisions attack at the same time they will win the battle, otherwise they will loose. Therefore, neither division attacks unless it is certain that the other will also attack. The general of one division wishes to coordinate a simultaneous attack with the general of the other division, and they can only do so by communicating with the help of messengers. In particular, no general attacks unless he has received a message. Normally, it takes one hour for a messenger to cross the valley, but the messenger might get lost or be captured by the enemy. Assuming nothing bad happens, how long will it take to coordinate an attack?

Call an interpreted context (γ, π) a message-delivery context if

- the environment and possibly some of the agents have actions that are message-delivery actions and results in messages being delivered to agents,
- γ is a recording context, and
- the language includes the proposition *delivered* which is true if at least one message has been delivered.

A message-delivery system is a system of the form $I^{rep}(P, \gamma, \pi)$ for a message-delivery context (γ, π) and a joint protocol P that can be run in γ.

Since *delivered* is false at the beginning of a run, it cannot be common knowledge at time 0. Moreover, in an a.m.p. system the agents can never achieve common knowledge of *delivered*. In fact, we will show that *delivered* can not become common knowledge in any system with sufficiently unreliable message passing, where sufficiently unreliable means that an agent can not be sure that a message got delivered unless it receives a message saying so.

Let $\mathcal{I} = (\mathcal{R}, \pi)$ be a message-delivery system. Given a run $r \in \mathcal{R}$, write $d(r, m) = k$ if exactly k messages have been delivered in the first m rounds of r. We say that $\mathcal{R}$ displays umd, unbounded message delivery, if for all $(r, m) \in \mathcal{R}$ with $d(r, m) > 0$ there is a agent i and a run $r' \in \mathcal{R}$ such that for all $j \neq i$ and times $m' \leq m$

$$r'_j(m') = r_j(m') \qquad \text{and} \qquad d(r', m) < d(r, m)$$

That is, there is a message delivered to i in r by round m that is not delivered to i in r' by round m and no agent except i knows this.

Many systems displays umd. For instance, a.m.p. and a.r.m.p systems display umd. Say that a context γ displays umd if $R^{rep}(P, \gamma)$ displays umd for every protocol P that can be run in the context γ. Then, the context implicitly described in the coordinated attack story displays umd.

Theorem 7.1. *Let $\mathcal{I} = (\mathcal{R}, \pi)$ be a message-delivery system such that $\mathcal{R}$ displays umd. Let G be a group of two or more agents. Then,*

$$\mathcal{I} \models \neg C_G \mathit{delivered}$$

Proof. The proof is by induction on $d(r, m)$. For $d(r, 0) = 0$, $(\mathcal{I}, r, 0) \models \neg C_G \mathit{delivered}$ obviously holds. Let $d(r, m) = k + 1$ and assume that the claim is true for all (r', m') with $d(r', m') \leq k$. Since $\mathcal{R}$ displays umd, there is a run r' such that $d(r', m) < d(r, m)$. Let i be the agent given by the umd definition. Then, $(\mathcal{I}, r', m) \models \neg C_G \mathit{delivered}$ by the induction hypothesis and $r'_j(m) = r_j(m)$ for an agent $j \neq i$. Thus, (r, m) is G-reachable from (r', m) and $(\mathcal{I}, r, m) \models \neg C_G \mathit{delivered}$. $\square$

We return to the coordinated attack problem. Call an interpreted context (γ, π) ca-compatible if it is a message-delivery context such that two of the agents are the generals A and B, and for $i \in \{A, B\}$, one of i's actions is attack_i. In addition, we require that there are propositions $\mathit{attacked}_i$ that are true at points where attack_i has been performed in the past. Let

$$\mathit{attacking}_i = \neg \mathit{attacked}_i \wedge \bigcirc \mathit{attacked}_i$$

and let $\mathit{attack} = \mathit{attacking}_A \wedge \mathit{attacking}_B$. Furthermore, let the specification σ^{ca} consist of all ca-compatible interpreted systems $\mathcal{I}$ such that

1. $\mathcal{I} \models \mathit{attacking}_A \Leftrightarrow \mathit{attacking}_B$,

2. $\mathcal{I} \models \neg delivered \Rightarrow \neg attack$, and

3. $(\mathcal{I}, r, m) \models attack$ for at least one point $(r, m) \in \mathcal{I}$.

A protocol P is a protocol for coordinated attack in a ca-compatible interpreted context (γ, π) if P satisfies σ^{ca} in (γ, π).

Theorem 7.2. *Let (γ, π) be a ca-compatible interpreted context and let P be a deterministic protocol. If $\mathcal{I} = I^{rep}(P, \gamma, \pi)$ satisfies σ^{ca} and $G = \{A, B\}$, then*

$$\mathcal{I} \models attack \Rightarrow C_G attack$$

Proof. Suppose $(\mathcal{I}, r, m) \models attack$ and $(\mathcal{I}, r, m) \models attacking_A \wedge attacking_B$. Then, $P_A\big(r_A(m)\big) = attack_A$. For any point (r', m') such that $r_A(m) = r'_A(m')$, $P_A\big(r_A(m)\big) = P_A\big(r'_A(m')\big) = attack_A$ since P is deterministic. Hence, $(\mathcal{I}, r, m) \models attacking_A$ and as $\mathcal{I}$ satisfies σ^{ca}, $(\mathcal{I}, r, m) \models attacking_B$. It follows that $(\mathcal{I}, r, m) \models K_A attack$. By a similar argument we get $(\mathcal{I}, r, m) \models K_B attack$ and therefore $(\mathcal{I}, r, m) \models E_G attack$.

Thus, $(\mathcal{I}, r, m) \models attack \Rightarrow E_G attack$ and the induction rule implies $(\mathcal{I}, r, m) \models attack \Rightarrow C_G attack$. ☐

Corollary 7.3. *Let (γ, π) be a ca-compatible interpreted context and let P be a deterministic protocol. If $\mathcal{I} = I^{rep}(P, \gamma, \pi)$ satisfies σ^{ca} and $G = \{A, B\}$, then*

$$\mathcal{I} \models attack \Rightarrow C_G delivered$$

Corollary 7.4. *If (γ, π) is a ca-compatible interpreted context such that γ displays umd, then there is no deterministic protocol P that satisfies σ^{ca} in (γ, π).*

A coordinated attack is therefore not possible.

7.2 Agreeing to disagree

An agent bases its actions on its knowledge, and its knowledge is determined by the states the agent considers possible. Thus, we want the actions to be based on the information set of the agent rather than on its local state.

Given a local state s_i for agent i in a system $\mathcal{R}$, let the information set $IS_i(s_i, \mathcal{R})$ be the set of points $(r, m) \in \mathcal{R}$ such that $r_i(m) = s_i$. If $\mathcal{I} = (\mathcal{R}, \pi)$, then we set $IS_i(s_i, \mathcal{I}) = IS_i(s_i, \mathcal{R})$.

For a set of global states $\mathcal{G}$, let S be the set of points over $\mathcal{G}$. Define a decision function for a process i to be a function from the subsets of S to

ACT_i, the actions of i. Decision functions are more general than protocols. If D is a decision function for i over the system $\mathcal{R}$, then we say that a protocol P_i is compatible with D in $\mathcal{R}$ if

$$P_i(s_i) = D\big(IS_i(s_i, \mathcal{R})\big)$$

for all $s_i \in L_i$.

We consider situations where all processes use the same decision function. Assume that ACT is the set of actions of the processes.

Definition 7.5. A joint protocol $P = (P_1, \dots, P_n)$ implements a decision function D in context γ if P_i is compatible with D in $R^{rep}(P, \gamma)$ for $i = 1, \dots, n$.

Call (γ, π^{ag}) an interpreted context for agreement if

- the processes actions are taken from the set ACT,
- γ is a recording context,
- for each action a, there is a primitive proposition $\mathit{perf}_i(a)$, and
- $\pi^{ag}(s)\big(\mathit{perf}_i(a)\big)$ is true in a state s if the action a was performed by i as recorded by the state of the environment.

Define $act_i(a) = \neg \mathit{perf}_i(a) \wedge \bigcirc \mathit{perf}_i(a)$.

Definition 7.6. A decision function D is union-consistent if for every action a and for every collection $T_1, T_2, \dots$ of pairwise disjoint subsets of S

$$D(T_j) = a \qquad \text{implies} \qquad D\Big(\bigcup_{j=1}^{\infty} T_j\Big) = a$$

Theorem 7.7. *Suppose $\mathcal{I} = I^{rep}(P, \gamma, \pi^{ag})$, where P is a joint protocol and (γ, π^{ag}) is an interpreted context for agreement. If P implements some union-consistent decision function in context γ, then for all actions $a, b \in ACT$,*

$$\mathcal{I} \models C_{\{1,2\}}\big(act_1(a) \wedge act_2(b)\big) \qquad \textit{implies} \qquad a = b$$

Proof. Assume that P implements a union-consistent decision function D in the context γ, and assume that $(\mathcal{I}, r, m) \models C\big(act_1(a) \wedge act_2(b)\big)$. Let S' consist of all points that are $\{1,2\}$-reachable from (r, m). Suppose $(r', m') \in S$ and $r'_1(m') = s_1$. Then, $(r'', m'') \in S'$ if $(r'', m'') \sim_1 (r', m')$ and hence $IS_1(s_1, \mathcal{I}) \subseteq S'$. In addition, if $s_1 \neq s'_1$, then $IS_1(s_1, \mathcal{I}) \cap IS_1(s'_1, \mathcal{I}) = \emptyset$. Thus, S' is a disjoint union of sets of the form $IS_1(s_1, \mathcal{I})$.

Since $(\mathcal{I}, r, m) \models C\big(act_1(a)\big)$, $(\mathcal{I}, r', m') \models act_1(a)$ and consequently $P_1(s_1) = a$. As P implements D, also $D\big(IS_1(s_1, \mathcal{I})\big) = a$. Finally, by the union-consistency of D, $D(S') = a$.

A similar argument shows $D(S') = b$. Hence, $a = b$. $\square$

It follows that if two agents follow the same union-consistent decision function, then they cannot have common knowledge of the fact that they take different actions, i.e., they cannot agree to disagree.

7.3 Simultaneous Byzantine agreement

There are n generals of which up to t might be traitors. Each general initially want to either attack or retreat. After having sent messages to each other according to some common protocol, all loyal generals must come to an agreement of whether to attack or retreat. A general that is a traitor may at any time fail to follow the protocol. In addition, we demand that if all generals have the same initial preferences, then they will agree upon this. In the simultaneous Byzantine agreement problem we require that the generals agree simultaneously.

An interpreted context is ba-compatible if

- for each process $i = 1, \ldots, n$, $decide_i(0)$ is an action of i representing a decision to retreat and $decide_i(1)$ is an action of i representing a decision to attack,
- each process has a $fail_i$ action which makes the process faulty,
- the actions of the environment includes actions of the form $(a_{e_1}, \ldots, a_{e_n})$ where a_{e_i} describes which messages are sent by i or describes that process i fails,
- γ is a recording context,
- the initial state of each process contains a variable $x_i \in \{0, 1\}$ which describes the initial preference of the process to attack or retreat, and the local state of the process records whether it has performed a $decide_i(x)$ action,
- the state of the environment includes the tuple $(x_1, \ldots, x_n)$ containing the initial preferences of the processes, and
- the language includes propositions $decided_i(x)$, $decided_N(x)$, and $\exists x$ for $x = 0, 1$ such that

- $decided_i(x)$ is true at a global state if the action $decide_i(x)$ was performed in any previous round,
- $decided_N(x)$ is true if each non-faulty process has performed the action $decide_i(x)$, and
- $\exists x$ is true if $x_i = x$ for some process i.

In addition, we let

$$deciding_N(x) = \neg decided_N(x) \wedge \bigcirc decided_N(x)$$

A ba-compatible interpreted system is of the form $I^{rep}(P, \gamma, \pi)$, where P is a protocol and (γ, π) is a ba-compatible interpreted context.

The specification σ^{sba} of simultaneous Byzantine agreement, SBA, is run-based. The ba-compatible interpreted system $\mathcal{I}$ satisfies σ^{sba} if every run r in $\mathcal{I}$ satisfy the conditions

- every process that is non-faulty in r performs exactly one $decide_i(x)$ action in r,
- the non-faulty processes decide on the same value,
- if all processes have the same initial preference, then they decide on that, and
- the non-faulty processes decide simultaneously, i.e., in the same round.

Let P be a protocol for SBA in a ba-compatible interpreted context (γ, π) if P satisfies σ^{sba} in (γ, π). If $r \in I^{rep}(P, \gamma, \pi) = \mathcal{I}$, then P attains SBA in k rounds in r if $(\mathcal{I}, r, k) \models deiciding_N$. The protocol P attains SBA in k round if P attains SBA in k rounds in all runs in $I^{rep}(P, \gamma, \pi)$.

Just as in the coordinated attack problem, attaining SBA requires common knowledge. However, only the non-faulty processes need to attain common knowledge and the set of non-faulty processes changes from point to point. A non-rigid set $\mathcal{S}$ of a system $\mathcal{R}$ is a function $\mathcal{S}(r, m)$ from the points of the system to the subsets of the processes. We write $i \in \mathcal{S}$ when i is in the non-rigid set $\mathcal{S}$. For a point (r, m), $i \in \mathcal{S}$ is true if $i \in \mathcal{S}(r, m)$.

Given a non-rigid set $\mathcal{S}$ and a process i, let $B^i_{\mathcal{S}}\varphi$ be an abbreviation for $K_i(i \in \mathcal{S} \Rightarrow \varphi)$. Then,

$$(\mathcal{I}, r, m) \models B^i_{\mathcal{S}}\varphi \qquad \text{iff} \qquad (\mathcal{I}, r', m') \models \varphi$$

for all (r', m') such that $r_i(m) = r'_i(m')$ and $i \in \mathcal{S}(r', m')$. Furthermore, define

$$E_{\mathcal{S}}\varphi = \bigwedge_{i \in \mathcal{S}} B^i_{\mathcal{S}}\varphi$$

and

$$C_{\mathcal{S}}\varphi = \bigwedge_{k=1}^{\infty} E_{\mathcal{S}}^{k}\varphi$$

We note that if $\mathcal{S}(r,m) = G$ for all points (r,m), then $C_{\mathcal{S}} = C_G$.

A point (r',m') is $\mathcal{S}$-reachable from a point (r,m) in k steps if there are points $(r_0,m_0),\ldots,(r_k,m_k)$ such that $(r,m) = (r_0,m_0)$, $(r',m') = (r_k,m_k)$, and there exists $i \in \mathcal{S}(r_l,m_l) \cap \mathcal{S}(r_{l+1},m_{l+1})$ such that $(r_l,m_l) \sim_i (r_{l+1},m_{l+1})$ for $l = 0,\ldots,k-1$.

Lemma 7.8. *$(\mathcal{I},r,m) \models C_{\mathcal{S}}\varphi$ if and only if $(\mathcal{I},r',m') \models \varphi$ for all points (r',m') that are $\mathcal{S}$-reachable from (r,m).*

Now, $B_{\mathcal{S}}^{i}$ satisfies all axioms of S5 except the knowledge axiom $B_{\mathcal{S}}^{i}\varphi \Rightarrow \varphi$, which is only true if $i \in \mathcal{S}$. Similarly, $C_{\mathcal{S}}$ satisfies all properties of common knowledge in S5 except possibly $C_{\mathcal{S}}\varphi \Rightarrow \varphi$. Moreover, $C_{\mathcal{S}}\varphi \Rightarrow \varphi$ is also valid if $\mathcal{S}(r,m) \neq \emptyset$ for all (r,m). In addition,

$$\models i \in \mathcal{S} \Rightarrow (B_{\mathcal{S}}^{i} C_{\mathcal{S}}\varphi \Leftrightarrow C_{\mathcal{S}}\varphi)$$

holds.

One can also define a non-rigid version of distributed knowledge:

$$(\mathcal{I},r,m) \models D_{\mathcal{S}}\varphi \qquad \text{iff} \qquad (\mathcal{I},r,m) \models D_G(G \subseteq \mathcal{S} \Rightarrow \varphi) \quad \text{for } G = \mathcal{S}(r,m)$$

Let N be a non-rigid set such that $N(r,m)$ consists of all non-faulty processes.

Theorem 7.9. *Let (γ,π) be a ba-compatible interpreted context and let P be a deterministic protocol. If $\mathcal{I} = I^{rep}(P,\gamma,\pi)$ satisfies σ^{sba}, then*

$$\mathcal{I} \models \mathit{deciding}_N(x) \Rightarrow C_N\big(\mathit{deciding}_N(x)\big)$$

Proof. Similar to Proposition 7.2. ∎

Corollary 7.10. *Let (γ,π) be a ba-compatible interpreted context and let P be a deterministic protocol. If $\mathcal{I} = I^{rep}(P,\gamma,\pi)$ satisfies σ^{sba}, then*

$$\mathcal{I} \models \mathit{deciding}_N(x) \Rightarrow C_N \exists x$$

7.4 Attaining SBA

Let $decided_i = decided_i(0) \vee decided_i(1)$. Informally, the process i follows the program

case of

  if $\neg decided_i \wedge B_N^i C_N \exists 0$ **do** $decide_i(0)$

  if $\neg decided_i \wedge \neg B_N^i C_N \exists 0 \wedge B_N^i C_N \exists 1$ **do** $decide_i(1)$

  $\vdots$

end case

where the if-statements not shown do not contain any $decide_i(x)$ actions. One can prove that there are deterministic protocols that attains SBA in $t+1$ rounds, where t is the number of traitors, if $n > 3t$, and that this is the best possible. A particular protocol that achieves this is the full information protocol: in every round every process sends its local state to all the other processes. We omit all proofs.

8 Knowledge and time

Time is characterized by the following axiom system

T1 $\bigcirc\varphi \wedge \bigcirc(\varphi \Rightarrow \psi) \Rightarrow \bigcirc\psi$

T2 $\bigcirc\neg\varphi \Leftrightarrow \neg\bigcirc\varphi$

T3 $\varphi\,\mathcal{U}\,\psi \Leftrightarrow \psi \vee \big(\varphi \wedge (\varphi\,\mathcal{U}\,\psi)\big)$

RT1 From φ infer $\bigcirc\varphi$

RT2 From $\varphi' \Rightarrow \neg\psi \wedge \bigcirc\varphi'$ infer $\varphi' \Rightarrow \neg(\varphi\,\mathcal{U}\,\psi)$

Let $\mathcal{L}_n^U$ be the language $\mathcal{L}_n$ augmented by the modal operators $\mathcal{U}$ and $\bigcirc$. Similarly, let $\mathcal{L}_n^{CU}$ be and augmentation of $\mathcal{L}_n^C$. Write $\mathcal{C}_n$ for the class of all interpreted systems.

Theorem 8.1. *For all formulas $\varphi, \psi \in \mathcal{L}_n^{CU}$ and for all interpreted systems $\mathcal{I} \in \mathcal{C}_n$*

1. $\mathcal{I} \models \bigcirc\varphi \wedge \bigcirc(\varphi \Rightarrow \psi) \Rightarrow \bigcirc\psi$

2. $\mathcal{I} \models \bigcirc\neg\psi \Leftrightarrow \neg\bigcirc\varphi$

3. $\mathcal{I} \models \varphi\,\mathcal{U}\,\psi \Leftrightarrow \psi \vee \big(\varphi \wedge (\varphi\,\mathcal{U}\,\psi)\big)$

4. If $\mathcal{I} \models \varphi$, *then* $\mathcal{I} \models \bigcirc\varphi$

5. If $\mathcal{I} \models \varphi' \Rightarrow \neg\psi \wedge \bigcirc\varphi'$, *then* $\mathcal{I} \models \varphi' \Rightarrow \neg(\varphi\,\mathcal{U}\,\psi)$

Proof. We prove 5. Take $\mathcal{I} \in \mathcal{C}_n$ and assume that $\mathcal{I} \models \varphi' \Rightarrow \neg\psi \wedge \bigcirc\varphi'$. By induction on k one can show $\mathcal{I} \models \varphi' \Rightarrow \bigcirc^k\varphi'$, from which follows $\mathcal{I} \models \varphi' \Rightarrow \Box\varphi'$. Since $\mathcal{I} \models \varphi' \Rightarrow \neg\psi$, we have $\mathcal{I} \models \varphi' \Rightarrow \Box\neg\psi$. Thus, ψ is never true later than φ'. Hence, $\mathcal{I} \models \varphi' \Rightarrow \neg(\varphi\,\mathcal{U}\,\psi)$. $\Box$

Let $\mathrm{S5}_n^U$ ($\mathrm{S5}_n^{CU}$) be the axiom system $\mathrm{S5}_n$ ($\mathrm{S5}_n^C$) to which has been added T1, T2, T3, RT1, and RT2.

Theorem 8.2. $\mathrm{S5}_n^U$ *(*$\mathrm{S5}_n^{CU}$*) is a sound and complete axiomatization of the language* $\mathcal{L}_n^U$ *(*$\mathcal{L}_n^{CU}$*) with respect to* $\mathcal{C}_n$.

Thus, if we consider all interpreted systems, then there is no interaction between knowledge and time.

Let $\mathcal{C}_n^{pr}$, $\mathcal{C}_n^{sync}$, and $\mathcal{C}_n^{pr,sync}$ be the interpreted systems where respectively perfect recall, synchrony, and perfect recall and synchrony hold.

Theorem 8.3. $\mathrm{S5}_n^U$ *(*$\mathrm{S5}_n^{CU}$*) is a sound and complete axiomatization of the language* $\mathcal{L}_n^U$ *(*$\mathcal{L}_n^{CU}$*) with respect to* $\mathcal{C}_n^{sync}$.

In systems with perfect recall time and knowledge do interact. The formula $K_i\varphi \Rightarrow \Box K_i\varphi$ is not valid in general, but it is valid for stable formulas, formulas that once true remains true. Since a stable formula φ satisfies $\varphi \Rightarrow \Box\varphi$ and since $\Box\varphi \Rightarrow \varphi$ is valid, we see that φ is equivalent to $\Box\varphi$. This suggests the axiom

KT1 $K_i\Box\varphi \Rightarrow \Box K_i\varphi$

which is valid in $\mathcal{C}_n^{pr}$. In synchronous systems a stronger axiom holds

KT2 $K_i \bigcirc \varphi \Rightarrow \bigcirc K_i\varphi$

Theorem 8.4. $\mathrm{S5}_n^U + KT2$ *is a sound and complete axiomatization of the language* $\mathcal{L}_n^U$ *with respect to* $\mathcal{C}_n^{pr,sync}$.

The axiom system $\mathrm{S5}_n^U + KT1$ is not complete with respect to $\mathcal{C}_n^{pr}$. One need a more complicated axiom in place of KT1:

$$K_i\varphi_1 \wedge K_i\neg\big((K_i\varphi_1 \vee K_i\varphi_3)\,\mathcal{U}\,\neg\varphi_2\big) \Rightarrow \bigcirc(K_i\varphi_3 \Rightarrow K_i\varphi_2)$$

If we add common knowledge, then there is no axiom system containing a finite number of axiom schemas and rules that is both sound and complete with respect to $\mathcal{C}_n^{pr}$ or $\mathcal{C}_n^{pr,sync}$.