Theoretical Computer Science 63 (1989) 63-90
North-Holland

COHESION OF OBJECT HISTORIES

Seymour GINSBURG*
Computer Science Department, University of Southern California, Los Angeles, CA 90089-0782, U.S.A.

Chang-jie TANG**
Computer Science Department, Chengdu, Sichuan, People's Republic of China

Communicated by M. Nivat
Received May 1987

Abstract. In an earlier paper, one of the authors introduced a record-based, algebraically-oriented, event-driven model for describing historical data for objects (here called "object histories"). The major construct in the model is a computation-tuple sequence scheme (CSS) which specifies the set of all possible "valid" object histories for the same type of object. The current paper considers the problem of combining the global information residing in a number of object histories in a distributed system. A suggested solution is in the form of an operation called "cohesion", which is the analogue for object histories of join for relational databases.

The basic question considered in this paper is the following: Given two sets $\mathcal{S}_1$ and $\mathcal{S}_2$ of object histories described by CSS T_1 and T_2, does there exist a CSS which describes the cohesion of $\mathcal{S}_1$ and $\mathcal{S}_2$? The answer is shown to be yes by constructing a specific CSS (called the "cohesion" of T_1 and T_2) from T_1 and T_2. The cohesion operation also turns out to be a useful tool for establishing some subsidiary results.

Introduction

In [2], a record-based, algebraically-oriented, event-driven model was introduced for describing historical data with computation for objects (called "object histories"). The major construct in the model is a computation-tuple sequence scheme (abbreviated CSS) which specifies the set of all possible "valid" object histories for the object of interest. The study of object histories was continued in a sequence of articles [1, 3, 4]. Essentially, all the work done so far has dealt with a single-site location. The question arises: How does one form a new object history which combines the global information residing in a number of object histories in a distributed system? The purpose of this paper is to suggest an answer by presenting a new operation called "cohesion", which is the analogue for object histories of join for relational databases.

* This author was supported in part by the National Science Foundation under Grant DCR-8318752.
** Research done at the University of Southern California while on leave from Sichuan University.

The basic question considered in this paper is the following: Given two sets $\mathcal{S}_1$ and $\mathcal{S}_2$ of object histories described by CSS T_1 and T_2 respectively, does there exist a CSS which describes the cohesion of $\mathcal{S}_1$ and $\mathcal{S}_2$? The answer is shown to be yes by constructing a specific CSS (called the "cohesion" of T_1 and T_2) from T_1 and T_2. The cohesion operation also turns out to be a useful tool for establishing some subsidiary results.

The paper itself is divided into four sections. The first reviews the object history model. Section 2 introduces the cohesion operation and presents a number of elementary results. Section 3 treats the basic question raised earlier. The final section shows the preservation of some properties of CSS under the cohesion operation.

1. Preliminaries

In this section, we present the model of object history introduced in [2]. The reader is referred to [2] for a more detailed discussion, together with motivational examples.

Informally, an object history is a historical record of an object. (Here, each object stands for an individual "thing" or "entity", such as a specific person's checking account, a specific company's sales record of an item, etc.) An object history is a sequence of occurrences, each occurrence consisting of some input data and, possibly, some calculation. (For example, in a checking-account history, one occurrence might be, in part, the amount to be deposited or withdrawn, together with the computation of the new balance and new daily minimum balance.) In the model, each object history is represented as a sequence of tuples (over the same attributes), called a "computation-tuple sequence." A CSS is a construct which defines the set of all possible "valid" computation-tuple sequences. (For example, a CSS for objects of the type "checking account" specifies the set of all possible "valid" individual checking-account histories.) A CSS consists of

(Δ1) a set of attributes, partitioned into state, input and evaluation attributes, according to their roles;

(Δ2) functions which calculate values for state and evaluation attributes;

(Δ3) semantic constraints whose satisfaction is to hold uniformly throughout a computation-tuple sequence; and

(Δ4) a set of specific computation-tuple sequences of some bounded length with which to start a valid computation-tuple sequence until all states and evaluation functions can be applied.

Turning to a formal treatment, Dom_∞ is an infinite set of elements (called *domain* values) and U_∞ is an infinite set of symbols (called *attributes*). For each A in U_∞, $\mathrm{Dom}(A)$ (called the *domain* of A) is a subset of Dom_∞ of at least two elements. All attributes occurring are assumed to be elements of U_∞. The symbols A, B and C (possibly subscripted) denote attributes and U, V and W (possibly subscripted or primed) denote nonempty finite sets of attributes.

Consistent with (Δ1), we shall assume the following.

Attribute Assumption. $U_\infty = S_\infty \cup I_\infty \cup E_\infty$, where S_∞, I_∞ and E_∞ are pairwise disjoint infinite sets (of *state* attributes, *input* attributes and *evaluation* attributes respectively). Furthermore, $\leq_\infty$ is a total order over U_∞ such that $A \leq_\infty B \leq_\infty C$ for each A in S_∞, B in I_∞ and C in E_∞.

Let X be a finite nonempty subset of U_∞ and $A_1, \ldots, A_n$ the listing of the elements of X according to $\leq_\infty$. Then $\langle X \rangle$ denotes the sequence $A_1 \ldots A_n$ and $\text{Dom}(\langle X \rangle)$ the cartesian product $\text{Dom}(A_1) \times \cdots \times \text{Dom}(A_n)$. For $i \geq 2$, $\langle X | A_i \rangle$ denotes the prefix $A_1 \ldots A_{i-1}$. (A *prefix* of a sequence $p_1 \ldots p_m$ is a subsequence of the form $p_1 \ldots p_i$ for some i, $1 \leq i \leq m$.)

We are now ready to formalize the notions of occurrence and sequence of occurrences as used earlier in this section. (Instead of "occurrence" and "sequence of occurrences" we shall use the terms "computation tuple" and "computation-tuple sequence".)

Definition. A *computation tuple* over $\langle U \rangle$ is an element in $\text{Dom}(\langle U \rangle)$. *A computation-tuple sequence* over $\langle U \rangle$ is a finite nonempty sequence of computation tuples over $\langle U \rangle$. The set of all computation-tuple sequences over $\langle U \rangle$ is denoted by $\text{SEQ}(\langle U \rangle)$.

Unless otherwise stated, u, v and w, possibly subscripted or primed, always represent computation tuples. Similarly, $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$ always represent computation-tuple sequences.

To formalize (Δ1) and (Δ2), we have the following definition.

Definition. A *computation scheme* (abbreviated CS) over $\langle U \rangle$ is a quintuple $\mathscr{C} = (\langle S \rangle, \langle I \rangle, \langle E \rangle, \mathscr{E}, \mathscr{F})$, where

(1) $S = S_\infty \cap U \neq \emptyset$, $I = I_\infty \cap U \neq \emptyset$ and $E = E_\infty \cap U$;

(2) $\mathscr{E} = \{e_C | C \text{ in } E, e_C$ a partial function (called an *evaluation* function) from $\text{Dom}(\langle U \rangle)^{\rho_C} \times \text{Dom}(\langle U | C \rangle)$ into $\text{Dom}(C)$ for some nonnegative integer $\rho_C\}$; and

(3) $\mathscr{F} = \{f_A | A \text{ in } S, f_A$ a partial function (called a *state* function) from $\text{Dom}(\langle U \rangle)$ into $\text{Dom}(A)\}$.

The integer ρ_C is called the *rank* of e_C; and $\rho(\mathscr{C}) = \max\{\rho_C, 1 | e_C \text{ in } \mathscr{E}\}$ is the *rank* of $\mathscr{C}$.

Intuitively, the rank of a computation scheme is the minimum number of previous computation tuples on which each computation tuple computationally depends.

Note that $\langle U \rangle = \langle S \rangle \langle I \rangle \langle E \rangle$.

Example 1.1. Consider the sales manager's record for a special souvenir, call it Sam Eagle, sold by a Los Angeles novelty company during (and after) the 1984 Olympic games. For simplicity, suppose this company has two retail outlets. The sale manager is responsible for

(1) collecting daily information on

- B_1: the amount ordered by outlet 1,
- B_2: the amount ordered by outlet 2,
- B_3: the price (in dollars) per item; and

(2) reporting to the warehouse manager about

- C_1: the (daily) total number ordered,
- C_2: the cost of C_1.

Using A (= "DATE") as a state attribute, B_1, B_2, B_3 as input attributes, and C_1, C_2 as evaluation attributes, the computation scheme is

$$\mathscr{C}_1 = (\langle A\rangle, \langle B_1B_2B_3\rangle, \langle C_1C_2\rangle, \{e_{1C_1}, e_{1C_2}\}, \{f_{1A}\})$$

described as follows (with $\langle U\rangle = \langle A\rangle\langle B_1B_2B_3\rangle\langle C_1C_2\rangle$):

(A) The domains of the attributes are the obvious ones.

(B) e_{1C_1} and e_{1C_2} are the functions from $\mathrm{Dom}(\langle U|C_1\rangle)$ to $\mathrm{Dom}(C_1)$ and $\mathrm{Dom}(\langle U|C_2\rangle)$ to $\mathrm{Dom}(C_2)$ respectively defined for each u in $\mathrm{Dom}(\langle U\rangle)$ by[1]

$$e_{1C_1}(u[\langle U|C_1\rangle]) = u(B_1) + u(B_2) \quad \text{and} \quad e_{1C_2}(u[\langle U|C_2\rangle]) = u(B_3)u(C_1).$$

(C) f_{1A} is the function from $\mathrm{Dom}(\langle U\rangle)$ to $\mathrm{Dom}(A)$ defined for each u in $\mathrm{SEQ}(\langle U\rangle)$ by

$$f_{1A}(u) = \text{"next date after } u(A)\text{"}.$$

The purpose of a computation scheme is to select those computation-tuple sequences whose values for the state and evaluation attributes are ultimately determined by the corresponding state and evaluation functions. More formally, we have the following notation.

Notation. Let $\mathscr{C} = (\langle S\rangle, \langle I\rangle, \langle E\rangle, \mathscr{E}, \mathscr{F})$ be a CS over $\langle U\rangle$. For each A in S and $\emptyset \neq S' \subseteq S$, let

$$\mathrm{VSEQ}(f_A) = \{u_1 \ldots u_m \mid m \geq 1,\ u_h(A) = f_A(u_{h-1}) \text{ for each } h,\ 2 \leq h \leq m\}$$

and

$$\mathrm{VSEQ}(\{f_A \mid A \text{ in } S'\}) = \bigcap_{A \text{ in } S'} \mathrm{VSEQ}(f_A).$$

For each C in E and $\emptyset \neq E' \subseteq E$, let

$$\mathrm{VSEQ}(e_C) = \{u_1 \ldots u_m \mid m \geq 1,\ u_h(C) = e_C(u_{h-\rho_C}, \ldots, u_{h-1}, u_h[\langle U|C\rangle]) \text{ for each } h,\ \rho_C < h \leq m\}$$

[1] Let $\langle U\rangle = A_1 \ldots A_n$ and $\langle V\rangle$ be a subsequence of $\langle U\rangle$. For each computation tuple u over $\langle U\rangle$, $u[\langle V\rangle]$ is the computation tuple v over $\langle V\rangle$ defined by $v(A) = u(A)$ for each A in V. $u[\langle V\rangle]$ is frequently written as $\Pi_V(u)$. For each $\mathbf{u} = u_1 \ldots u_m$ in $\mathrm{SEQ}(\langle U\rangle)$, $\Pi_V(\mathbf{u}) = \Pi_V(u_1) \ldots \Pi_V(u_m)$. For each $\mathscr{S} \subseteq \mathrm{SEQ}(\langle U\rangle)$, $\Pi_V(\mathscr{S}) = \{\Pi_V(\mathbf{u}) \mid \mathbf{u} \text{ in } S\}$.

and

$$\text{VSEQ}(\{e_C \mid C \text{ in } E'\}) = \bigcap_{C \text{ in } E'} \text{VSEQ}(e_C).$$

Let VSEQ(∅) = SEQ(⟨*U*⟩) and VSEQ($\mathscr{C}$) = VSEQ($\mathscr{E}$) ∩ VSEQ($\mathscr{F}$).

Clearly,

$$\text{VSEQ}(\mathscr{E}) = \{u_1 \ldots u_m \mid m \geq 1,\ u_h(C) = e_C(u_{h-\rho_C}, \ldots, u_{h-1}, u_h[\langle U|C\rangle]) \text{ for each } C \text{ in } E \text{ and each } h,\ \rho_C < h \leq m\}$$

and

$$\text{VSEQ}(\mathscr{F}) = \{u_1 \ldots u_m \mid m \geq 1,\ u_h(A) = f_A(u_{h-1}) \text{ for each } A \text{ in } S \text{ and each } h,\ 2 \leq h \leq m\}.$$

Obviously, u is in VSEQ(f_A) iff each interval u_1u_2 of u is in VSEQ(f_A); and u is in VSEQ(e_C) iff each interval $u_1 \ldots u_{\rho_C+1}$ of u is in VSEQ(e_C). (An *interval* of a sequence $p_1 \ldots p_m$ is a subsequence of the form $p_i \ldots p_j$ for each i and j, $1 \leq i \leq j \leq m$.) Also, VSEQ($\mathscr{C}$) is an interval-closed set. Note that VSEQ(f_A) contains all computation tuples, and VSEQ(e_C) all computation-tuple sequences of length at most ρ_C. In effect, VSEQ(g), g a function, consists of all computation-tuple sequences which do not "contradict" the functioning of g.

Example 1.1 (*continued*). From the definitions, it follows that

$$\text{VSEQ}(e_{1C_1}) = \{u_1 \ldots u_m \mid m \geq 1,\ u_i(C_1) = u_i(B_1) + u_i(B_2) \text{ for all } i, 1 \leq i \leq m\},$$

$$\text{VSEQ}(e_{1C_2}) = \{u_1 \ldots u_m \mid m \geq 1,\ u_i(C_2) = u_i(B_3)u_i(C_1) \text{ for all } i,\ 1 \leq i \leq m\},$$

$$\text{VSEQ}(f_{1A}) = \{u_1 \ldots u_m \mid m \geq 1,\ u_{i+1}(A) = \text{"next date after } u_i(A)\text{" for all } i,\ 1 \leq i \leq m-1\} \text{ and}$$

$$\text{VSEQ}(\mathscr{C}_1) = \{u_1 \ldots u_m \mid m \geq 1,\ u_i(C_1) = u_i(B_1) + u_i(B_2),\ u_i(C_2) = u_i(B_3)u_i(C_1),\ u_{j+1}(A) = \text{"next date after } u_j(A)\text{" for all } i \text{ and } j,\ 1 \leq i \leq m \text{ and } 1 \leq j \leq m-1\}.$$

Turning to constraints, i.e. (Δ3), we have this definition.

Definition. A *constraint* σ over SEQ(⟨*U*⟩) is a mapping over SEQ(⟨*U*⟩) which assigns to each u in SEQ(⟨*U*⟩) a value of "true" or "false". If $\sigma(u)$ = true, then u is said to *satisfy* σ. For each set Σ of constraints over SEQ(⟨*U*⟩), the set {u in SEQ(⟨*U*⟩) | u satifies each σ in Σ} is denoted by VSEQ(Σ).

Note that VSEQ(Σ) = SEQ(⟨*U*⟩) if $\Sigma = \emptyset$.

We shall usually define a constraint σ by just specifying VSEQ(σ).

The concept of a constraint given above is too general to be mathematically tractable. We shall restrict our constraints to a special class called "uniform". These are characterized by the fact that satisfaction holds uniformly throughout a computation-tuple sequence, i.e. holds in every interval of a computation-tuple sequence. (Most constraints encountered in real life are of this type.)

Definition. A constraint σ over SEQ($\langle U\rangle$) is *uniform* if VSEQ(σ) is interval closed, i.e. if $\boldsymbol{u}$ is in VSEQ(σ), then so is every interval of $\boldsymbol{u}$.

Clearly, VSEQ(Σ) is interval closed for each set Σ of uniform constraints.

Example 1.1 (*continued*). The set Σ_1 of constraints is empty.

The last concept needed for a computation-tuple sequence scheme is the "initialization". (See (Δ4)).

Definition. Given a CS $\mathscr{C}$ over $\langle U\rangle$ and a finite set Σ of uniform constraints over SEQ($\langle U\rangle$), an *initialization* (*with respect to* $\mathscr{C}$ *and* Σ) is any prefix-closed subset $\mathscr{I}$ of[2]

$$\{\boldsymbol{u} \text{ in VSEQ}(\mathscr{C}) \cap \text{VSEQ}(\Sigma) \mid |\boldsymbol{u}| \leq \rho(\mathscr{C})\}.$$

Given an initialization $\mathscr{I}$, let VSEQ($\mathscr{I}$) denote the set

$$\mathscr{I} \cup \{\boldsymbol{u} \text{ in SEQ}(\langle U\rangle) \mid \boldsymbol{u} = \boldsymbol{u}_1\boldsymbol{u}_2 \text{ for some } \boldsymbol{u}_1 \text{ in } \mathscr{I} \text{ of length } \rho(\mathscr{C})\}.$$

Clearly, each VSEQ($\mathscr{I}$) is prefix closed but not necessarily interval closed.

Example 1.1 (*continued*). The initialization $\mathscr{I}_1$ is

$$\{(\text{date}, b_1, b_2, b_3, b_1+b_2, b_3(b_1+b_2)) \mid \text{date in Dom}(A),\ b_i \text{ in Dom}(B_i),\ 1 \leq i \leq 3\}.$$

We are now ready to define the fundamental notion of computation-tuple sequence scheme.

Definition. A *computation-tuple sequence scheme* (*CSS*) over ($\langle S\rangle, \langle I\rangle, \langle E\rangle$) (abbreviated "over $\langle U\rangle$", with $\langle U\rangle = \langle S\rangle\langle I\rangle\langle E\rangle$) is a triple $T = (\mathscr{C}, \Sigma, \mathscr{I})$, where
(1) $\mathscr{C}$ is a computation scheme over $\langle U\rangle$;
(2) Σ is a finite set of uniform constraints over SEQ($\langle U\rangle$); and
(3) $\mathscr{I}$ is an initialization with respect to $\mathscr{C}$ and Σ.
Let $\rho(T)$, called the *rank* of T, be $\rho(\mathscr{C})$.

[2] $|\boldsymbol{u}|$ denotes the length of $\boldsymbol{u}$.

A CSS determines valid computation-tuple sequences as follows.

Definition. For each CSS $T=((\langle S\rangle, \langle I\rangle, \langle E\rangle, \mathscr{E}, \mathscr{F}), \Sigma, \mathscr{I})$, let

$$\mathrm{VSEQ}(T)=\mathrm{VSEQ}(\mathscr{E})\cap\mathrm{VSEQ}(\mathscr{F})\cap\mathrm{VSEQ}(\Sigma)\cap\mathrm{VSEQ}(\mathscr{I}).$$

A computation-tuple sequence is said to be *valid* (*for* T) if it is in VSEQ(T).

Thus, a computation-tuple sequence is valid if it
(i) is "consistent" with $\mathscr{C}$,
(ii) satisfies each constraint in Σ, and
(iii) is either in the initialization or its prefix, of length $\rho(\mathscr{C})$, is in the initialization.

Example 1.1 (*continued*). A valid computation-tuple sequence $u_1u_2u_3$ for $T_1=(\mathscr{C}_1, \Sigma_1, \mathscr{I}_1)$ is given in Table 1.

Since both VSEQ($\mathscr{C}$) and VSEQ(Σ) are interval closed and VSEQ($\mathscr{I}$) is prefix closed, VSEQ(T) is prefix closed. However, VSEQ(T) is not necessarily interval closed.

Note that if $T_i=((\langle S_i\rangle, \langle I_i\rangle, \langle E_i\rangle, \mathscr{E}_i, \mathscr{F}_i), \Sigma_i, \mathscr{I}_i)$ for $i=1,2$ and $\mathrm{VSEQ}(T_1)\subseteq\mathrm{VSEQ}(T_2)$, then $\langle S_1\rangle=\langle S_2\rangle$, $\langle I_1\rangle=\langle I_2\rangle$ and $\langle E_1\rangle=\langle E_2\rangle$ by the Attribute Assumption.

Table 1. Sales manager's record.

	$\langle S_1\rangle$	$\langle I_1\rangle$			$\langle E_1\rangle$	
	A Date	B_1 Amount ordered by outlet 1	B_2 Amount ordered by outlet 2	B_3 Price per item	C_1 Total number ordered	C_2 Cost of C_1
u_1	7-26-84	3,000	5,000	5	8,000	40,000
u_2	7-27-84	3,000	6,000	6	9,000	54,000
u_3	7-28-84	4,000	7,000	6	11,000	66,000

2. Basic concepts

As mentioned in the Introduction, the purpose of this paper is to introduce and study the operation of cohesion, which is the analogue for computation-tuple sequences of natural join for database relations. In this section, we define the concept and that of rank-r minimum representation. We also note some elementary properties about cohesion and establish the existence of rank-r minimum representation.

To motivate our central concept, we present the following example.

Example 2.1. Expanding on Example 1.1, the warehouse manager is responsible for

(1) collecting information on

- B_3: price (in dollars) per item,
- B_4: (daily) amount delivered (for simplicity, we assume that the warehouse delivers to each outlet the exact number ordered),
- B_5: (daily) amount received (of Sam Eagle souvenirs from the manufacturer); and

(2) reporting about

- C_2: cost of C_1 (=daily total number ordered),
- C_3: number (of souvenirs) available.

A CSS for the records of the warehouse manager is

$$T_2 = ((\langle A\rangle, \langle B_3B_4B_5\rangle, \langle C_2C_3\rangle, \{e_{2C_2}, e_{2C_3}\}, \{f_{2A}\}), \Sigma_2, \mathscr{I}_2)$$

over $\langle V\rangle = \langle A\rangle\langle B_3B_4B_5\rangle\langle C_2C_3\rangle$ described as follows:

(A) The domains of the attributes are the obvious ones.

(B) e_{2C_2} is the function from $\text{Dom}(\langle V|C_2\rangle)$ to $\text{Dom}(C_2)$ defined for each v in $\text{SEQ}(\langle V\rangle)$ by

$$e_{2C_2}(v[\langle V|C_2\rangle]) = v(B_3)v(B_4),$$

and e_{2C_3} is the function from $\text{Dom}(\langle V\rangle)\times\text{Dom}(\langle V|C_3\rangle)$ to $\text{Dom}(C_3)$ defined for each v_1v_2 in $\text{SEQ}(\langle V\rangle)$ by

$$e_{2C_3}(v_1, v_2[\langle V|C_3\rangle]) = v_1(C_3) + v_2(B_5) - v_2(B_4).$$

(C) f_{2A} is the function from $\text{Dom}(\langle V\rangle)$ to $\text{Dom}(A)$ defined for each v in $\text{SEQ}(\langle V\rangle)$ by $f_{2A}(v)$ = "the next date after $v(A)$".

(D) $\Sigma_2 = \emptyset$.

(E) $\mathscr{I}_2 = \{(a, b_3, b_4, b_5, b_3b_4, b_5 - b_4) \mid a \text{ in Dom}(A),\ b_i \text{ in Dom}(B_i) \text{ for } 3 \leq i \leq 5\}$.

A valid computation-tuple sequence $v_1v_2v_3$ for T_2 is given in Table 2.

The global information in Tables 1 and 2, call it a valid record for the general manager, is given in Table 3.

The operation (called "cohesion") which merges the information in Tables 1 and 2 to yield that in Table 3 is the central concept of the present paper. It is formalized as follows.

Table 2. Warehouse manager's record.

	$\langle S_2\rangle$	$\langle I_2\rangle$			$\langle E_2\rangle$	
	A Date	B_3 Price per item	B_4 Amount delivered	B_5 Amount received	C_2 Cost of C_1	C_3 Number available
v_1	7-26-84	5	8,000	20,000	40,000	12,000
v_2	7-27-84	6	9,000	10,000	54,000	13,000
v_3	7-28-84	6	11,000	10,000	66,000	12,000

Table 3.

	$\langle S_1S_2\rangle$	$\langle I_1I_2\rangle$					$\langle E_1E_2\rangle$		
	A	B_1	B_2	B_3	B_4	B_5	C_1	C_2	C_3
w_1	7-26-84	3,000	5,000	5	8,000	20,000	8,000	40,000	12,000
w_2	7-27-84	3,000	6,000	6	9,000	10,000	9,000	54,000	13,000
w_3	7-28-84	4,000	7,000	6	11,000	10,000	11,000	66,000	12,000

Definition. Given $\langle U\rangle = \langle S_1I_1E_1\rangle$ and $\langle V\rangle = \langle S_2I_2E_2\rangle$, the *cohesion* of u in SEQ($\langle U\rangle$) and v in SEQ($\langle V\rangle$), denoted $u © v$, is

(1) the computation-tuple sequence w in[3] SEQ($\langle S_1S_2I_1I_2E_1E_2\rangle$) such that $\Pi_{S_1I_1E_1}(w) = u$ and $\Pi_{S_2I_2E_2}(w) = v$ if $\Pi_A(u) = \Pi_A(v)$ for each A in $(S_1I_1E_1) \cap (S_2I_2E_2)$, and

(2) undefined, denoted $\emptyset$, otherwise.

The *cohesion* of $\mathcal{S}_1 \subseteq$ SEQ($\langle U\rangle$) and $\mathcal{S}_2 \subseteq$ SEQ($\langle V\rangle$), denoted $\mathcal{S}_1 © \mathcal{S}_2$, is the set $\{u © v \mid u \text{ in } \mathcal{S}_1, v \text{ in } \mathcal{S}_2\}$.

By the Attribute Assumption, cohesion is well defined.

It is readily seen that cohesion is commutative, associative and idempotent. Because of associativity, we may omit the grouping parentheses when dealing with cohesion of more than two items. Also, $\mathcal{S}_1 © \mathcal{S}_2 = \mathcal{S}_1 \cap \mathcal{S}_2$ if $\mathcal{S}_1 \subseteq$ SEQ($\langle U\rangle$) and $\mathcal{S}_2 \subseteq$ SEQ($\langle U\rangle$).

Since grouping parentheses may be omitted, we have the following notation.

Notation. For $n \geq 2$ and $i = 1, \ldots, n$, let u_i be in SEQ($\langle U_i\rangle$) and $\mathcal{S}_i \subseteq$ SEQ($\langle U_i\rangle$). Then

$$\underset{1\leq i\leq n}{©} u_i = u_1 © \cdots © u_n$$

and

$$\underset{1\leq i\leq n}{©} \mathcal{S}_i = \mathcal{S}_1 © \cdots © \mathcal{S}_n.$$

A number of easily proved, frequently used, properties of cohesion with respect to projection are summarized (without proof) in the next result.

Proposition 2.2. *For* $i = 1, 2$ *let* $\mathcal{S}_i \subseteq$ SEQ($\langle U_i\rangle$) *and* u_i *be in* $\mathcal{S}_i$. *Then*

(a) $\Pi_{U_j}(u_1 © u_2) = u_j$ *for* $j = 1, 2$ *if* $u_1 © u_2 \neq \emptyset$;

(b) $\Pi_{U_j}(\mathcal{S}_1 © \mathcal{S}_2) \subseteq \mathcal{S}_j$ *for* $j = 1, 2$;

(c) *for each* u *in* SEQ($\langle U_1U_2\rangle$), $u = \Pi_{U_1}(u) © \Pi_{U_2}(u)$; *and*

(d) *for each* $\mathcal{S}$ *in* SEQ($\langle U_1U_2\rangle$), $\mathcal{S} \subseteq \Pi_{U_1}(\mathcal{S}) © \Pi_{U_2}(\mathcal{S})$.

We shall also have occasion to use the following readily established result on the distributivity of intersection with respect to cohesion (proof omitted).

[3] As usual, if X and Y are sets of attributes, then XY is the union of X and Y.

Proposition 2.3. *For $n \geq 2$ and $i = 1, \ldots, n$ let $\mathcal{S}_i \subseteq \mathrm{SEQ}(\langle U \rangle)$ and $\mathcal{T}_i \subseteq \mathrm{SEQ}(\langle V \rangle)$. Then $\bigcap_{i=1}^n (\mathcal{S}_i \copyright \mathcal{T}_i) = (\bigcap_{i=1}^n \mathcal{S}_i) \copyright (\bigcap_{i=1}^n \mathcal{T}_i)$.*

Although not used in the sequel, we note that

$$\left(\bigcup_{i=1}^{m} \mathcal{S}_i\right) \copyright \left(\bigcup_{j=1}^{n} \mathcal{T}_j\right) = \bigcup_{i,j} (\mathcal{S}_i \copyright \mathcal{T}_j)$$

for all $\mathcal{S}_i \subseteq \mathrm{SEQ}(\langle U \rangle)$ and all $\mathcal{T}_j \subseteq \mathrm{SEQ}(\langle V \rangle)$.

For the remainder of this section, we concentrate on the notion of a "minimum representation" (with respect to cohesion). To motivate this concept, suppose a distributed system has a CSS T_1 over $\langle U_1 \rangle$ at site 1 and a CSS T_2 over $\langle U_2 \rangle$ at site 2. Furthermore, suppose we wish to compute $u_1 \copyright u_2$ (for a given u_1 in $\mathrm{VSEQ}(T_1)$ and u_2 in $\mathrm{VSEQ}(T_2)$) via a channel in which communication is relatively expensive. One approach to reducing the communication cost is to seek T_1' over $\langle U_1 \rangle$ and T_2' over $\langle U_2 \rangle$ such that $\mathrm{VSEQ}(T_1')$ and $\mathrm{VSEQ}(T_2')$ are "minimum" (under the containment relation) sets satisfying

$$\mathrm{VSEQ}(T_1') \copyright \mathrm{VSEQ}(T_2') = \mathrm{VSEQ}(T_1) \copyright \mathrm{VSEQ}(T_2).$$

(That is,

(a) $\mathrm{VSEQ}(T_1') \copyright \mathrm{VSEQ}(T_2') = \mathrm{VSEQ}(T_1) \copyright \mathrm{VSEQ}(T_2)$, and

(b) $\mathrm{VSEQ}(T_i') \subseteq \mathrm{VSEQ}(T_i'')$ for each CSS T_i'' over $\langle U_i \rangle$, $1 \leq i \leq 2$, such that $\mathrm{VSEQ}(T_1') \copyright \mathrm{VSEQ}(T_2') = \mathrm{VSEQ}(T_1'') \copyright \mathrm{VSEQ}(T_2'')$.)

Indeed, suppose such a T_1' and T_2' exist. Then we first determine whether or not u_i ($i = 1, 2$) is in $\mathrm{VSEQ}(T_i')$. If u_i is not in $\mathrm{VSEQ}(T_i')$, then there is no point in considering u_i for cohesion purposes. Unfortunately, as we now show, such a T_1' and T_2' need not exist.

Example 2.4.[4] Let $T_1 = ((\langle A_1 \rangle, \langle B \rangle, \langle C \rangle, \{e_C\}, \{f_{A_1}\}), \{\sigma_1\}, \mathcal{I}_1)$ and $T_2 = ((\langle A_2 \rangle, \langle B \rangle, \emptyset, \emptyset, \{f_{A_2}\}), \{\sigma_2\}, \mathcal{I}_2)$ over $\langle U \rangle = \langle A_1 \rangle \langle B \rangle \langle C \rangle$ and $\langle V \rangle = \langle A_2 \rangle \langle B \rangle$ respectively be defined as follows:

(A) The domain of each attribute is the integers.

(B) e_C is the function on $\mathrm{Dom}(\langle U \rangle) \times \mathrm{Dom}(\langle A_1 B \rangle)$ defined for each $u_1 u_2$ in $\mathrm{SEQ}(\langle U \rangle)$ by $e_C(u_1, u_2[\langle A_1 B \rangle]) = 0$.

(C) f_{A_1} and f_{A_2} are the mappings over $\mathrm{Dom}(\langle U \rangle)$ and $\mathrm{Dom}(\langle V \rangle)$ respectively defined for each u in $\mathrm{Dom}(\langle U \rangle)$ and v in $\mathrm{Dom}(\langle V \rangle)$ by $f_{A_1}(u) = 0$, $f_{A_2}(v) = v(A_2) - 1$ if $v(A_2) > 1$ and $f_{A_2}(v) = v(B) + 1$ otherwise.

(D) Let $\mathrm{VSEQ}(\sigma_1) = \mathrm{SEQ}(\langle U \rangle)$. Let $s_1, s_2, \ldots$ be the infinite sequence of elements in $\mathrm{Dom}(\langle V \rangle)$ where $s_1 = (1, 1)$ and for each $i \geq 1$, $s_{i+1}(A_2) = f_{A_2}(s_i)$, $s_{i+1}(B) = s_i(B)$ if $s_{i+1}(A_2) \neq 1$, and $s_{i+1}(B) = s_i(B) + 1$ if $s_{i+1}(A_2) = 1$. Thus, the sequence begins with

$$(1, 1), (2, 2), (1, 2), (3, 3), (2, 3), (1, 3), (4, 4), (3, 4).$$

Let $\mathrm{VSEQ}(\sigma_2) = \{s_i s_{i+1} \ldots s_j \mid \text{all } i \text{ and } j,\ 1 \leq i \leq j\}$.

[4] We wish to thank Stephen Kurtzman for providing this example, thereby replacing a much more complicated one originally given by us.

(E) $\mathscr{I}_1 = \{(0, n, 0) \mid n \text{ in Dom}(A_1)\}$ and $\mathscr{I}_2 = \{(n, n) \mid n > 0\}$. Clearly,

$$\text{VSEQ}(T_1) = \{(0, i_1, 0)\ldots(0, i_n, 0) \mid n \geq 1, \text{ each } i_j \text{ in Dom}(A_1)\} \quad \text{and}$$

$$\text{VSEQ}(T_2) = \{s_i \ldots s_j \mid s_i = (n, n) \text{ for some } n \geq 1, j \geq i\}.$$

For each l, let t_l be the tuple $(0, s_l(A_2), s_l(B), 0)$ in $\text{Dom}(\langle UV\rangle)$. Then

$$\text{VSEQ}(T_1) \copyright \text{VSEQ}(T_2) = \{t_i \ldots t_j \mid t_i = (0, l, l, 0) \text{ for some } l \geq 1, j \geq i\}.$$

We now show that (T_1, T_2) has no minimum representation. Indeed, suppose (T'_1, T'_2) is a minimum representation of (T_1, T_2). Thus,

$$\text{VSEQ}(T'_1) \copyright \text{VSEQ}(T'_2) = \text{VSEQ}(T_1) \copyright \text{VSEQ}(T_2)$$

and

$$\text{VSEQ}(T'_2) \subseteq \text{VSEQ}(T_2).$$

Let

$$T'_1 = ((\langle A_1\rangle, \langle B\rangle, \langle C\rangle, \{e'_C\}, \{f'_{A_1}\}), \{\sigma'_1\}, \{\mathscr{I}'_1\}),$$

let $\rho_C = \gamma$ be the rank of e'_C and ρ the rank of T'_1. Note that $\rho = 1$ if $\gamma = 0$, and $\rho = \gamma$ if $\gamma > 0$. Let

$$T''_1 = ((\langle A_1\rangle, \langle B\rangle, \langle C\rangle, \{e''_C\}, \{f'_{A_1}\}), \{\sigma'_1\}, \mathscr{I}''_1)$$

be the CSS of rank $\gamma + 2 > \rho$, where

(F) $$e''_C(u_1, \ldots, u_{\gamma+2}, u_{\gamma+3}[A_1, B]) = e'_C(u_3, \ldots, u_{\gamma+2}, u_{\gamma+3}[A_1, B])$$

for all $u_1 \ldots u_{\gamma+3}$ in $\text{SEQ}(\langle U\rangle)$,

and

(G) $$\mathscr{I}''_1 = \{u \text{ in VSEQ}(T'_1) \mid |u| \leq \gamma + 2,\ u \copyright v \neq \emptyset \quad \text{for some } v \text{ in VSEQ}(T'_2)\}.$$

Clearly, $\mathscr{I}''_1$ is prefix closed and $\text{VSEQ}(T''_1) \subseteq \text{VSEQ}(T'_1)$. Also, it is straightforward to see that

$$\text{VSEQ}(T''_1) \copyright \text{VSEQ}(T'_2) = \text{VSEQ}(T'_1) \copyright \text{VSEQ}(T'_2).$$

To show that (T_1, T_2) has no minimum representation it therefore suffices to prove that $\text{VSEQ}(T''_1) \subsetneq \text{VSEQ}(T'_1)$, i.e. there exists some u in $\text{VSEQ}(T'_1) - \text{VSEQ}(T''_1)$.

Let

$$w_1 = (0, \gamma+2, \gamma+2, 0)(0, \gamma+1, \gamma+2, 0)\ldots(0, 1, \gamma+2, 0)(0, \gamma+3, \gamma+3, 0).$$

Obviously, w_1 is in $\text{VSEQ}(T'_1) \copyright \text{VSEQ}(T'_2)$. Let

(1) $u_1 = \Pi_U(w_1) = (0, \gamma+2, 0)^{\gamma+2}(0, \gamma+3, 0)$.

By Proposition 2.2(a),

(2) u_1 is in $\text{VSEQ}(T'_1)$.

Let $u = (0, \gamma+2, 0)^{\rho}(0, \gamma+3, 0)$. Since ρ is the rank of T'_1, u is a suffix of u_1, and since $\text{VSEQ}(e'_C)$, $\text{VSEQ}(f_{A_1})$ and $\text{VSEQ}(\sigma'_1)$ are interval closed, it follows that u is in VSEQ(e'_C) $\cap$ VSEQ(f_{A_1}) $\cap$ VSEQ(σ'_1). Since $\rho=\gamma$ or $\rho=\gamma+1$, $(0, \gamma+2, 0)^\rho$ is in $\mathscr{I}'_1$ by (1) and (2). Hence, u is in VSEQ($\mathscr{I}'_1$) and therefore in VSEQ(T'_1). Suppose u is in VSEQ(T''_1). Since $|u|=\rho+1\leq\gamma+2$, u is in $\mathscr{I}''_1$. By (G), there exists a v in VSEQ(T'_2) such that u © $v\neq\emptyset$. Now

$$v=(\gamma+2, \gamma+2)(\gamma+1, \gamma+2)\ldots(\gamma+2-\rho, \gamma+2)$$

is the only sequence in VSEQ(T'_2) $\subseteq$ VSEQ(T_2) of length $\rho+1$ which begins with a tuple v such that $v(B)=\gamma+2$. However, u © $v=\emptyset$. Thus u is not in $\mathscr{I}''_1$, i.e. u is in VSEQ(T'_1) $-$ VSEQ(T''_1) as was desired.

The above example shows that if the ranks of the various T''_i are allowed to be arbitrarily large, then no minimum representation need exist. However, it will be shown in Proposition 2.5 below that if the ranks of the T''_i are required to be bounded, then there is a minimum representation. (We will not consider the problem of presenting reasonable sufficient conditions for a minimum representation to exist when the rank of the various T''_i are allowed to be arbitrarily large.) In preparation for that result we have the following definition.

Definition. For $n\geq 2$ and each i, $1\leq i\leq n$, let T_i be a CSS over $\langle U_i\rangle$. Let r be a positive integer. An n-tuple $(T'_1,\ldots,T'_n)$ of CSS is said to be a *rank-r minimum representation* of $(T_1,\ldots,T_n)$ (with respect to *cohesion*) if

(a) $©_{1\leq i\leq n}$ VSEQ(T'_i) $=$ $©_{1\leq i\leq n}$ VSEQ(T_i), and

(b) VSEQ(T'_i) $\subseteq$ VSEQ(T''_i) for each i, $1\leq i\leq n$, for each CSS T''_i over $\langle U_i\rangle$, of rank at most r, such that

$$\underset{1\leq i\leq n}{©} \mathrm{VSEQ}(T''_i)=\underset{1\leq i\leq n}{©} \mathrm{VSEQ}(T_i).$$

$(T_1,\ldots,T_n)$ is said to be a rank-r minimum representation if it is a rank-r minimum representation of itself.

Obviously, if $(T'_1,\ldots,T'_n)$ and $(T''_1,\ldots,T''_n)$ are both rank-r minimum representations of $(T_1,\ldots,T_n)$, with each T'_j and T''_k of rank at most r, then VSEQ(T'_i) $=$ VSEQ(T''_i) for all i, $1\leq i\leq n$.

We now show that a rank-r minimum representation always exists.

Proposition 2.5.[5] *For all* CSS $T_1,\ldots,T_n$ *and all* $r\geq\max\{\rho(T_i)\,|\,1\leq i\leq n\}$, *a rank-r minimum representation* $(T'_1,\ldots,T'_n)$ of $(T_1,\ldots,T_n)$ *exists. Furthermore,* $\rho(T'_i)\leq r$ *for each i.*

Proof. For $n=1$, there is nothing to prove. Thus, assume $n\geq 2$. For each i, let $T_i=((\langle S_i\rangle, \langle I_i\rangle, \langle E_i\rangle, \mathscr{E}_i, \mathscr{F}_i), \Sigma_i, \mathscr{I}_i)$.

[5] The authors wish to thank Mr. Guozhu Dong for discussions leading to a clarification of the argument in Proposition 2.5.

(a) Suppose that for each i, either $\rho(T_i)=r$ or $E_i=\emptyset$. For each i, let $\langle U_i\rangle = \langle S_i\rangle\langle I_i\rangle\langle E_i\rangle$, $©_{j\neq i}$ VSEQ(T_j) be the cohesion of all VSEQ(T_j), $j\neq i$, and

$$T_i' = ((\langle S_i\rangle, \langle I_i\rangle, \langle E_i\rangle, \mathscr{E}_i, \mathscr{F}_i), \{\sigma_i\}, \mathscr{I}_i')$$

where

$$\text{VSEQ}(\sigma_i) = \text{Interval}^{6}\left(\left\{u \text{ in VSEQ}(T_i) \mid u \,©\, v \neq \emptyset \text{ for some } v \text{ in } \underset{j\neq i}{©} \text{VSEQ}(T_j)\right\}\right)$$

and

$$\mathscr{I}_i' = \left\{u \text{ in } \mathscr{I}_i \mid u \,©\, v \neq \emptyset \text{ for some } v \text{ in } \underset{j\neq i}{©} \text{VSEQ}(T_j)\right\}.$$

For each i, it is clear that VSEQ(σ_i) is interval closed (so that σ_i is a uniform constraint), $\mathscr{I}_i'$ is prefixed closed and $\rho(T_i')=\rho(T_i)\leq r$.

We now show that $(T_1',\ldots,T_n')$ is a rank-r minimum representation of $(T_1,\ldots,T_n)$. To do this, it is enough to prove that

(1) $©_{1\leq i\leq n}$ VSEQ(T_i') = $©_{1\leq i\leq n}$ VSEQ(T_i), and

(2) VSEQ$(T_i')\subseteq$ VSEQ(T_i'') for all i, $1\leq i\leq n$, and all CSS T_i'' over $\langle U_i\rangle$, of rank at most r, such that $©_{1\leq i\leq n}$ VSEQ(T_i'') = $©_{1\leq i\leq n}$ VSEQ(T_i').

Consider (1). Obviously, the left side is a subset of the right. We now examine the reverse inclusion. For each i, let

$$\mathscr{S}_i = \left\{u \text{ in VSEQ}(T_i) \mid u \,©\, v \neq \emptyset \text{ for some } v \text{ in } \underset{j\neq i}{©} \text{VSEQ}(T_j)\right\}.$$

Clearly, $©_{1\leq i\leq n}$ VSEQ(T_i) = $©_{1\leq i\leq n}\,\mathscr{S}_i$. Let i, $1\leq i\leq n$, be fixed. We first show that

(3) $\mathscr{S}_i\subseteq$ VSEQ(T_i').

To this end, let u be in $\mathscr{S}_i$. Then

(4) u is in VSEQ(T_i) and

(5) there exists some v in $©_{j\neq i}$ VSEQ(T_j) such that $u \,©\, v\neq\emptyset$.

By (4), u in VSEQ$(\mathscr{E}_i)\cap$VSEQ$(\mathscr{F}_i)$. By (5), u is in VSEQ(σ_i). To establish (3), it thus suffices to show that

(6) u is in VSEQ$(\mathscr{I}_i')$.

Two cases arise.

Case (α): suppose $|u|\leq r$. By (4), u is in VSEQ$(\mathscr{I}_i)$ and hence in $\mathscr{I}_i$. Combining this with (5), it follows that u is in $\mathscr{I}_i'\subseteq$ VSEQ$(\mathscr{I}_i')$.

Case (β): suppose $|u|>r$. Let $u=u_1u_2$ and $v=v_1v_2$, where $|u_1|=|v_1|=r$. Since VSEQ(T) is closed under prefix for all CSS T, (4) and (5) hold when u and v are replaced by u_1 and v_1 respectively. By (α), u_1 is in $\mathscr{I}_i'$. Hence (6) holds, and (3) is proven.

[6] For each set $\mathscr{U}\subseteq$ SEQ$(\langle U\rangle)$, Interval$(\mathscr{U})=\{u' \mid u'$ an interval of some u in $\mathscr{U}\}$.

By (3), $\text{©}_{1\le i\le n}\mathcal{S}_i \subseteq \text{©}_{1\le i\le n}\text{VSEQ}(T_i')$. Then

$$\underset{1\le i\le n}{\text{©}}\ \text{VSEQ}(T_i) = \underset{1\le i\le n}{\text{©}}\ \mathcal{S}_i \subseteq \underset{1\le i\le n}{\text{©}}\ \text{VSEQ}(T_i')$$

as desired, so (1) holds.

Consider (2). For each i, $1\le i\le n$, let $T_i''=(((\langle S_i\rangle,\langle I_i\rangle,\langle E_i\rangle,\mathscr{C}_i'',\mathscr{F}_i''),\Sigma_i'',\mathscr{I}_i'')$ be CSS, of rank at most r, such that $\text{©}_{1\le i\le n}\text{VSEQ}(T_i'')=\text{©}_{1\le i\le n}\text{VSEQ}(T_i)$. Let i be arbitrary, $1\le i\le n$. If $\rho(T_i)=r$, then $\rho(T_i'')\le r=\rho(T_i')$. If $E_i=\emptyset$, then $\rho(T_i'')=\rho(T_i')=1$. In either case,

(7) $\rho(T_i'')\le\rho(T_i')$.

Now

$$\begin{aligned}
\mathscr{I}_i' &= \Big\{u \text{ in } \mathscr{I}_i \,\Big|\, u \text{ © } v\neq\emptyset \text{ for some } v \text{ in } \underset{j\neq i}{\text{©}}\ \text{VSEQ}(T_j)\Big\}\\
&\subseteq \Big\{u \text{ in VSEQ}(T_i) \,\Big|\, u \text{ © } v\neq\emptyset \text{ for some } v \text{ in } \underset{j\neq i}{\text{©}}\ \text{VSEQ}(T_j)\Big\}\\
&= \Big\{\Pi_{U_i}\Big(u_i \text{ © }\Big(\underset{j\neq i}{\text{©}}\ u_j\Big)\Big) \,\Big|\, u_i \text{ © }\Big(\underset{j\neq i}{\text{©}}\ u_j\Big)\neq\emptyset, \text{ each } u_k \text{ in VSEQ}(T_k),\\
&\qquad 1\le k\le n\Big\}, \quad \text{by Proposition 2.2(a)}\\
&= \Pi_{U_i}\Big(\underset{1\le j\le n}{\text{©}}\ \text{VSEQ}(T_j)\Big)\\
&= \Pi_{U_i}\Big(\underset{1\le j\le n}{\text{©}}\ \text{VSEQ}(T_j'')\Big) \quad \text{by assumption}\\
&\subseteq \text{VSEQ}(T_i''), \quad \text{by Proposition 2.2(b)}\\
&\subseteq \text{VSEQ}(\mathscr{I}_i'').
\end{aligned}$$

That is,

(8) $\mathscr{I}_i'\subseteq\text{VSEQ}(\mathscr{I}_i'')$.

Then

$$\begin{aligned}
\text{VSEQ}(\mathscr{I}_i') &= \mathscr{I}_i'\cup\{uw \mid u \text{ in } \mathscr{I}_i',\ |u|=\rho(T_i'),\ w \text{ in SEQ}(\langle U_i\rangle)\}\\
&\subseteq \text{VSEQ}(\mathscr{I}_i'')\cup\{uw \text{ in VSEQ}(\mathscr{I}_i''),\ |u|=\rho(T_i'),\ w \text{ in SEQ}(\langle U_i\rangle)\}\\
&\qquad \text{by (8)}\\
&\subseteq \text{VSEQ}(\mathscr{I}_i'')\cup\{xyw \mid x \text{ in } \mathscr{I}_i'',\ |x|=\rho(T_i''),\ yw \text{ in SEQ}(\langle U_i\rangle),\\
&\qquad y \text{ possibly empty}\} \quad \text{by (7)}\\
&\subseteq \text{VSEQ}(\mathscr{I}_i'')\cup\text{VSEQ}(\mathscr{I}_i'')\\
&= \text{VSEQ}(\mathscr{I}_i'').
\end{aligned}$$

That is,

(9) $\text{VSEQ}(\mathscr{I}_i')\subseteq\text{VSEQ}(\mathscr{I}_i'')$.

Clearly,

$$
\begin{aligned}
\mathrm{VSEQ}(\sigma_i) &= \mathrm{Interval}\Big(\Big\{\boldsymbol{u} \text{ in } \mathrm{VSEQ}(T_i) \,|\, \boldsymbol{u} © \boldsymbol{v} \neq \emptyset \\
&\qquad\qquad \text{for some } \boldsymbol{v} \text{ in } \underset{j\neq i}{©} \mathrm{VSEQ}(T_j)\Big\}\Big) \\
&= \mathrm{Interval}\Big(\Big\{\Pi_{U_i}(\boldsymbol{u}_i © \boldsymbol{v}) \,|\, \boldsymbol{u}_i © \boldsymbol{v} \neq \emptyset \\
&\qquad\qquad \boldsymbol{u}_i \text{ in } \mathrm{VSEQ}(T_i),\ \boldsymbol{v} \text{ in } \underset{j\neq i}{©} \mathrm{VSEQ}(T_j)\Big\}\Big) \\
&\qquad\qquad \text{by Proposition 2.2(a)} \\
&= \mathrm{Interval}(\Pi_{U_i}(©_{i\leq j\leq n} \mathrm{VSEQ}(T_j))) \\
&= \mathrm{Interval}(\Pi_{U_i}(©_{i\leq j\leq n} \mathrm{VSEQ}(T_j''))) \quad \text{by assumption} \\
&\subseteq \mathrm{Interval}(\mathrm{VSEQ}(T_i'')) \quad \text{by Proposition 2.2(b)} \\
&\subseteq \mathrm{Interval}(\mathrm{VSEQ}(\mathscr{E}_i'') \cap \mathrm{VSEQ}(\mathscr{F}_i'') \cap \mathrm{VSEQ}(\Sigma_i'')) \\
&\subseteq \mathrm{Interval}(\mathrm{VSEQ}(\mathscr{E}_i'')) \cap \mathrm{Interval}(\mathrm{VSEQ}(\mathscr{F}_i'')) \\
&\qquad\qquad \cap \mathrm{Interval}(\mathrm{VSEQ}(\Sigma_i'')) \\
&= \mathrm{VSEQ}(\mathscr{E}_i'') \cap \mathrm{VSEQ}(\mathscr{F}_i'') \cap \mathrm{VSEQ}(\Sigma_i''),
\end{aligned}
$$

since $\mathrm{VSEQ}(\mathscr{E}_i'')$, $\mathrm{VSEQ}(\mathscr{F}_i'')$ and $\mathrm{VSEQ}(\Sigma_i'')$ are interval closed. That is,

(10) $\mathrm{VSEQ}(\sigma_i) \subseteq \mathrm{VSEQ}(\mathscr{E}_i'') \cap \mathrm{VSEQ}(\mathscr{F}_i'') \cap \mathrm{VSEQ}(\Sigma_i'')$.

Then

$$
\begin{aligned}
\mathrm{VSEQ}(T_i') &\subseteq \mathrm{VSEQ}(\mathscr{I}_i') \cap \mathrm{VSEQ}(\sigma_i) \\
&\subseteq \mathrm{VSEQ}(\mathscr{I}_i'') \cap \mathrm{VSEQ}(\mathscr{E}_i'') \cap \mathrm{VSEQ}(\mathscr{F}_i'') \cap \mathrm{VSEQ}(\Sigma_i'') \\
&\qquad\qquad \text{by (9) and (10)} \\
&= \mathrm{VSEQ}(T_i''),
\end{aligned}
$$

as desired.

(b) Consider the general case. The following result was established in [3, Lemma 2.2]: "Let $T = (\mathscr{C}, \Sigma, \mathscr{I})$ be a CSS with at least one evaluation attribute. Then for each integer $r > \rho(T)$, there exists a CSS $T' = (\mathscr{C}', \Sigma, \mathscr{I}')$ such that $\rho(T') = r$ and $\mathrm{VSEQ}(T') = \mathrm{VSEQ}(T)$".[7] By this result, for each i, either

(i) there exists a T_i of rank r such that $\mathrm{VSEQ}(T_i) = \mathrm{VSEQ}(T_i)$ or

(ii) $E_i = \emptyset$, in which case, let $T_i = T_i$.

By case (a) above, there exists a rank-r minimum representation $(T_1', \ldots, T_n')$ of $(T_1, \ldots, T_n)$, with $\rho(T_i') \leq r$ for each i. Since $\mathrm{VSEQ}(T_i) = \mathrm{VSEQ}(T_i)$ for each i, $(T_1', \ldots, T_n')$ is a rank-r minimum representation of $(T_1, \ldots, T_n)$. □

[7] Indeed, let $T = (((\langle S\rangle, \langle I\rangle, \langle E\rangle, \mathscr{E}, \mathscr{F}), \Sigma, \mathscr{I})$. One such T' is $(((\langle S\rangle, \langle I\rangle, \langle E\rangle, \mathscr{E}', \mathscr{F}'), \Sigma, \mathscr{I}')$, where $\mathscr{I}' = \{\boldsymbol{u}$ in $\mathrm{VSEQ}(T) \,|\, |\boldsymbol{u}| \leq r\}$, $\mathscr{F}' = \mathscr{F}$ and $\mathscr{E}' = \{e_C' \,|\, C \text{ in } E\}$. Here, e_C' is the (partial) function from $\mathrm{Dom}(\langle U\rangle)^r \times \mathrm{Dom}(\langle U|C\rangle)$ into $\mathrm{Dom}(C)$ defined by

$$e_C'(v_1, \ldots, v_{r-\rho_C}, u_1, \ldots, u_{\rho_C}, u_{\rho_C+1}[\langle U|C\rangle]) = e_C(u_1, \ldots, u_{\rho_C}, u_{\rho_C+1}[\langle U|C\rangle])$$

for each $v_1 \ldots v_{r-\rho_C} u_1 \ldots u_{\rho_C+1}$ in $\mathrm{SEQ}(\langle U\rangle)$.

We now illustrate the rank-r minimum representation.

Example 2.6. Let T_1 and T_2 be the CSS, of rank 1 each, given in Examples 1.1 and 2.1 respectively. Let (T_1', T_2') be the rank-1 minimum representation of (T_1, T_2) as constructed in the argument of Proposition 2.5. Then

$$T_1' = ((\langle A\rangle, \langle B_1B_2B_3\rangle, \langle C_1C_2\rangle, \{e_{1C_1}, e_{1C_2}\}, \{f_{1A}\}), \{\sigma_1\}, \mathcal{I}_1')$$

and

$$T_2' = ((\langle A\rangle, \langle B_3B_4B_5\rangle, \langle C_2C_3\rangle, \{e_{2C_2}, e_{2C_3}\}, \{f_{2A}\}), \{\sigma_2\}, \mathcal{I}_2')$$

where

$$\text{VSEQ}(\sigma_1) = \text{Interval}(\{u \text{ in VSEQ}(T_1) \mid u \circledcirc v \neq \emptyset \text{ for some } v \text{ in VSEQ}(T_2)\}),$$

$$\text{VSEQ}(\sigma_2) = \text{Interval}(\{v \text{ in VSEQ}(T_2) \mid u \circledcirc v \neq \emptyset \text{ for some } u \text{ in VSEQ}(T_1)\}),$$

$$\mathcal{I}_1' = \{u \text{ in } \mathcal{I}_1 \mid u \circledcirc v \neq \emptyset \text{ for some } v \text{ in VSEQ}(T_2)\},$$

and

$$\mathcal{I}_2' = \{v \text{ in } \mathcal{I}_2 \mid u \circledcirc v \neq \emptyset \text{ for some } u \text{ in VSEQ}(T_1)\}.$$

It is readily seen that $\text{VSEQ}(T_1') = \text{VSEQ}(T_1)$ and $\text{VSEQ}(T_2') = \text{VSEQ}(T_2)$. Thus (T_1, T_2) is a rank-1 minimum representation of itself. On the other hand, suppose we modify the original T_2 so that $\Sigma_2 = \{\sigma\}$, where σ is the constraint defined by

$$\text{VSEQ}(\sigma) = \{v_1 \ldots v_n \text{ in VSEQ}(\langle U_2\rangle) \mid v_i(C_3) \leq 50{,}000 \text{ for each } i\}.$$

(This means that the warehouse can accommodate at most 50,000 "Sam Eagles" at one time.) Then (T_1, T_2) is not a rank-1 minimum representation. For example, $u = (8\text{-}1\text{-}84, 30\,000, 30000, 5, 60\,000, 300\,000)$ is in $\text{VSEQ}(T_1)$ but not in $\text{VSEQ}(T_1')$, i.e. there is no v in $\text{VSEQ}(T_2)$ such that $u \circledcirc v \neq \emptyset$.

3. Cohesion of CSS

As already noted, we are interested in the cohesion of $\text{VSEQ}(T_1)$ and $\text{VSEQ}(T_2)$. It is thus natural to ask: does there exist a CSS T_3 such that $\text{VSEQ}(T_3) = \text{VSEQ}(T_1) \circledcirc \text{VSEQ}(T_2)$? In this section we show that the answer is yes. Indeed, we shall "construct" a specific such T_3 and refer to its as the "cohesion" of T_1 and T_2.

We start by introducing some special symbolism.

Notation. (1) Let f_{1A} and f_{2A} be (state) functions from $\text{Dom}(\langle U\rangle)$ into $\text{Dom}(A)$. Then $f_{1A} \circledcirc f_{2A}$ denotes the function f_A from $\text{Dom}(\langle U\rangle)$ into $\text{Dom}(A)$ defined for each u in $\text{Dom}(\langle U\rangle)$ by $f_A(u) = f_{1A}(u)$ if $f_{1A}(u) = f_{2A}(u)$, and $f_A(u) = \emptyset$ otherwise.

(2) Let e_{1C} and e_{2C} be (evaluation) functions from $\text{Dom}(\langle U\rangle)^{\rho_1}\times\text{Dom}(\langle U|C\rangle)$ into $\text{Dom}(C)$ and from $\text{Dom}(\langle U\rangle)^{\rho_2}\times\text{Dom}(\langle U|C\rangle)$ into $\text{Dom}(C)$ respectively. Let $\rho=\max\{\rho_1,\rho_2\}$. Then $e_{1C} © e_{2C}$ denotes the function e_C from $\text{Dom}(\langle U\rangle)^{\rho}\times \text{Dom}(\langle U|C\rangle)$ into $\text{Dom}(C)$ defined for each $u_1\ldots u_{\rho+1}$ in $\text{SEQ}(\langle U\rangle)$ by

$$e_C(u_1,\ldots,u_\rho,u_{\rho+1}[\langle U|C\rangle])=e_{1C}(u_{\rho-\rho_1+1},\ldots,u_\rho,u_{\rho+1}[\langle U|C\rangle])$$

if

$$e_{1C}(u_{\rho-\rho_1+1},\ldots,u_\rho,u_{\rho+1}[\langle U|C\rangle])=e_{2C}(u_{\rho-\rho_2+1},\ldots,u_\rho,u_{\rho+1}[\langle U|C\rangle]),$$

and is undefined otherwise.

Thus, $f_{1A} © f_{2A}$ is the function representing exactly where f_{1A} and f_{2A} coincide. $e_{1C} © e_{2C}$ is the function representing exactly where e'_{1C} and e'_{2C} coincide, e'_{1C} and e'_{2C} being the functions e_{1C} and e_{2C} respectively converted to rank $\max\{\rho_1,\rho_2\}$.

It is easily seen that © is commutative and associative among state functions and among evaluation functions.

Notation. Let $S_1\subseteq S$, $I_1\subseteq I$, $E_1\subseteq E$, $\langle U\rangle=\langle S\rangle\langle I\rangle\langle E\rangle$ and $\langle V\rangle=\langle S_1\rangle\langle I_1\rangle\langle E_1\rangle$.

(1) Suppose A is in S_1 and f_A is a (state) function from $\text{Dom}(\langle V\rangle)$ into $\text{Dom}(A)$. Then f_A^U is the (state) function from $\text{Dom}(\langle U\rangle)$ to $\text{Dom}(A)$ defined for each u in $\text{SEQ}(\langle U\rangle)$ by $f_A^U(u)=f_A(\Pi_V(u))$.

(2) Suppose C is in E_1 and e_C is an (evaluation) function from $\text{Dom}(\langle V\rangle)^{\rho_C}\times \text{Dom}(\langle V|C\rangle)$ into $\text{Dom}(C)$. Then e_C^U is the (evaluation) function from $\text{Dom}(\langle U\rangle)^{\rho_C}\times \text{Dom}(\langle U|C\rangle)$ into $\text{Dom}(C)$ defined for each $u_1\ldots u_{\rho_C+1}$ in $\text{SEQ}(\langle U\rangle)$ by

$$e_C^U(u_1,\ldots,u_{\rho_C},u_{\rho_C+1}[\langle U|C\rangle])=e_C(\Pi_V(u_1),\ldots,\Pi_V(u_{\rho_C}),u_{\rho_C+1}[\langle V|C\rangle]).$$

Thus, f_A^U is the function obtained from f_A by extending the domain from $\langle V\rangle$ to $\langle U\rangle$ and then ignoring the effect of the added attributes. e_C^U is obtained from e_C in essentially the same manner.

We are now ready to present the central concept of the section.

Definition. Let $T_1=(((\langle S_1\rangle,\langle I_1\rangle,\langle E_1\rangle,\mathscr{E}_1,\mathscr{F}_1),\Sigma_1,\mathscr{I}_1)$, and $T_2=(((\langle S_2\rangle,\langle I_2\rangle,\langle E_2\rangle,\mathscr{E}_2,\mathscr{F}_2),\Sigma_2,\mathscr{I}_2)$ be CSS, $\langle W\rangle=\langle S_1S_2\rangle\langle I_1I_2\rangle\langle E_1E_2\rangle$ and $\rho=\max\{\rho(T_1),\rho(T_2)\}$. The *cohesion* of T_1 and T_2, denoted $T_1 © T_2$, is $(((\langle S_1S_2\rangle,\langle I_1I_2\rangle,\langle E_1E_2\rangle,\mathscr{E},\mathscr{F}),\{\sigma\},\mathscr{I})$, where

(a) $\mathscr{E}=\{e_C\,|\,e_C=e_{1C}^W,\ C\text{ in }E_1-E_2\}\cup\{e_C\,|\,e_C=e_{2C}^W,\ C\text{ in }E_2-E_1\}\cup$ $\{e_C\,|\,e_C=e_{1C}^W © e_{2C}^W,\ C\text{ in }E_1\cap E_2\}$;

(b) $\mathscr{F}=\{f_A\,|\,f_A=f_{1A}^W,\ A\text{ in }S_1-S_2\}\cup\{f_A\,|\,f_A=f_{2A}^W,\ A\text{ in }S_2-S_1\}\cup$ $\{f_A\,|\,f_A=f_{1A}^W © f_{2A}^W,\ A\text{ in }S_1\cap S_2\}$;

(c) $\text{VSEQ}(\sigma)=\text{VSEQ}(\Sigma_1) © \text{VSEQ}(\Sigma_2)$; and

(d) $\mathscr{I}=\{u © v\,|\,u\text{ in VSEQ}(T_1),\ v\text{ in VSEQ}(T_2),\ |u|\le\rho,\ |v|\le\rho\}$.

Clearly, T is a CSS over $\langle W\rangle$ of rank ρ; and $\mathscr{I}=\mathscr{I}_1 © \mathscr{I}_2$ if $\rho(T_1)=\rho(T_2)$. Also, the cohesion of CSS is associative, commutative and idempotent.

Example 3.1. For T_1 and T_2 as in Examples 1.1 and 2.1 respectively,

$$T_1 © T_2 = ((\langle A\rangle, \langle B_1B_2B_3B_4B_5\rangle, \langle C_1C_2C_3\rangle, \{e_{C_1}, e_{C_2}, e_{C_3}\}, \{f_A\}), \Sigma, \mathcal{I})$$

is the CSS over $\langle W\rangle = \langle A\rangle\langle B_1B_2B_3B_4B_5\rangle\langle C_1C_2C_3\rangle$ defined as follows:

(A) e_{C_1} is the function from Dom($\langle W|C_1\rangle$) to Dom(C_1) defined for each w in SEQ($\langle W\rangle$) by $e_{C_1}(w[\langle W|C_1\rangle]) = w(B_1)+w(B_2)$, e_{C_2} is the function from Dom($\langle W|C_2\rangle$) to Dom(C_2) defined for each w in SEQ($\langle W\rangle$) by $e_{C_2}(w[\langle W|C_2\rangle]) = w(B_3)w(C_1)$, and e_{C_3} is the evaluation function from Dom($\langle W\rangle$) × Dom($\langle W|C_3\rangle$) to Dom(C_3) defined for each w_1w_2 in SEQ($\langle W\rangle$) by $e_{C_3}(w_1, w_2[\langle W|C_3\rangle]) = w_1(C_3)+w_2(B_5)-w_2(B_4)$.

(B) f_A is the state function from Dom($\langle W\rangle$) to Dom(A) defined for each w in SEQ($\langle W\rangle$) by $f_A(w)$ = "the next date after $w(A)$".

(C) $\Sigma = \emptyset$.

(D) $\mathcal{I} = \{u © v | u$ in VESQ(T_1), v in VESQ(T_2)$\}$.

The main result about $T_1 © T_2$ is that VSEQ($T_1 © T_2$) = VSEQ(T_1) © VSEQ(T_2). To prove this, we first establish a lemma dealing with relationships between the various VSEQs of $T_1 © T_2$ and the corresponding VSEQs of T_1 and T_2.

Lemma 3.2. *For $i = 1, 2$ let $T_i = (((\langle S_i\rangle, \langle I_i\rangle, \langle E_i\rangle, \mathscr{E}_i, \mathscr{F}_i), \Sigma_i, \mathcal{I}_i)$, $T = T_1 © T_2 = (((\langle S_1S_2\rangle, \langle I_1I_2\rangle, \langle E_1E_2\rangle, \mathscr{E}, \mathscr{F}), \{\sigma\}, \mathcal{I})$ and $\mathcal{I}'_i = \{y$ in VSEQ(T_i)$\,|\, |y| \le \rho(T)\}$. Then*

(a) VSEQ($\mathscr{E}$) ∩ VSEQ($\mathcal{I}$) = (VSEQ($\mathscr{E}_1$) ∩ VSEQ($\mathcal{I}'_1$)) © (VSEQ($\mathscr{E}_2$) ∩ VSEQ($\mathcal{I}'_2$)),

(b) VSEQ($\mathscr{F}$) = VSEQ($\mathscr{F}_1$) © VSEQ($\mathscr{F}_2$),

(c) VSEQ(σ) = VSEQ(Σ_1) © VSEQ(Σ_2), *and*

(d) VSEQ($\mathcal{I}$) = VSEQ($\mathcal{I}'_1$) © VSEQ($\mathcal{I}'_2$).

Proof. It follows from the definition of cohesion of CSS that (c) holds and that

(*) $\quad \mathcal{I} = \mathcal{I}'_1 © \mathcal{I}'_2$.

Clearly, (d) follows from (*). Thus, only (a) and (b) need to be examined. Let $\langle U\rangle = \langle S_1I_1E_1\rangle$, $\langle V\rangle = \langle S_2I_2E_2\rangle$ and $\langle W\rangle = \langle UV\rangle$. Let $w = w_1 \ldots w_m$ be in SEQ($\langle W\rangle$), $\Pi_U(w) = u = u_1 \ldots u_m$ and $\Pi_V(w) = v = v_1 \ldots v_m$. By Proposition 2.2(c), $w = u © v$.

Turning to (a), it suffices to verify that

(1) $\quad$ VSEQ($\mathscr{E}$) ∩ VSEQ($\mathcal{I}$)
$\quad\quad \subseteq$ (VSEQ($\mathscr{E}_1$) ∩ VSEQ($\mathcal{I}'_1$)) © (VSEQ($\mathscr{E}_2$) ∩ VSEQ($\mathcal{I}'_2$))

and

(2) $\quad$ (VSEQ($\mathscr{E}_1$) ∩ VSEQ($\mathcal{I}'_1$)) © (VSEQ($\mathscr{E}_2$) ∩ VSEQ($\mathcal{I}'_2$))
$\quad\quad \subseteq$ VSEQ($\mathscr{E}$) ∩ VSEQ($\mathcal{I}$).

Consider (1). Suppose w is in the left side of (1). To show that w is in the right side of (1), it suffices (by symmetry and the fact that $w = u © v$) to prove that

(3) u is in VSEQ($\mathscr{E}_1$) ∩ VSEQ($\mathcal{I}'_1$).

Consider (3). Two cases arise:

Case (α): $m \leq \rho(T)$. Since w is in VSEQ($\mathcal{S}$) and $|w| = m \leq \rho(T)$, w is in $\mathcal{S} = \mathcal{S}_1' © \mathcal{S}_2'$. Thus $u = \Pi_U(w)$ is in

$$\begin{aligned} \Pi_U(\mathcal{S}_1' © \mathcal{S}_2') &\subseteq \mathcal{S}_1', && \text{by Proposition 2.2(b)} \\ &\subseteq \text{VSEQ}(T_1) && \text{by definition of } \mathcal{S}_1' \\ &\subseteq \text{VSEQ}(\mathcal{E}_1) \cap \text{VSEQ}(\mathcal{S}_1') \end{aligned}$$

as desired.

Case (β): $m > \rho(T)$. Since the left side of (1) is prefix closed,

(4) $u_1 \ldots u_{\rho(T)}$ is in $\mathcal{S}_1'$

by (α). Hence, u is in VSEQ($\mathcal{S}_1'$). Let $\mathcal{E} = \{e_C \mid C \text{ in } E_1E_2\}$, $\mathcal{E}_1 = \{e_{1C} \mid C \text{ in } E_1\}$ and $\mathcal{E}_2 = \{e_{2C} \mid C \text{ in } E_2\}$. To complete the argument for (3) it is enough to show that u is in VSEQ($\mathcal{E}_1$), i.e.

(5) $u_1 \ldots u_m$ is in VSEQ(e_{1C}) for each C in E_1.

Consider (5). Two subcases arise.

Subcase (β1): C is in $E_1 - E_2$. Let ρ_{1C} be the rank of e_{1C}. By the definition of cohesion of CSS, $e_C = e_{1C}^W$ and hence is of rank

$$\rho_{1C} < \rho_{1C} + 1 \leq \rho(T) + 1 \leq m.$$

Then for each i, $\rho_{1C} + 1 \leq i \leq m$,

$$u_i(C) = w_i(C)$$

since $\Pi_U(w_i) = u_i$

$$= e_C(w_{i-\rho_{1C}}, w_{i-\rho_{1C}+1}, \ldots, w_{i-1}, w_i[\langle W|C\rangle])$$

since w is in VSEQ($\mathcal{E}$) $\subseteq$ VSEQ(e_C)

$$= e_{1C}(u_{i-\rho_{1C}}, \ldots, u_{i-1}, u_i[\langle U|C\rangle])$$

since $e_C = e_{1C}^W$. Hence, u is in VSEQ(e_{1C}) as desired.

Subcase (β2): C is in $E_1 \cap E_2$. Let ρ_{1C} and ρ_{2C} be the ranks of e_{1C} and e_{2C} respectively. Then $e_C = e_{1C}^W © e_{2C}^W$ has rank $\rho_C = \max\{\rho_{1C}, \rho_{2C}\}$. By (4),

(6) $u_1 \ldots u_{\rho(T)}$ is in $\mathcal{S}_1' \subseteq \text{VSEQ}(T_1) \subseteq \text{VSEQ}(e_{1C})$.

Note that $\rho_C \leq \rho(T)$. For each i, $\rho(T) + 1 \leq i \leq m$,

$$(7) \qquad u_i = w_i(C),$$

since $\Pi_U(w_i) = u_i$

$$= e_C(w_{i-\rho_C}, \ldots, w_{i-1}, w_i[\langle W|C\rangle])$$

since w is in VSEQ(T) $\subseteq$ VSEQ(e_C) and e_C is of rank ρ_C

$$= e_{1C}(u_{i-\rho_{1C}}, \ldots, u_{i-1}, u_i[\langle W|C\rangle])$$

since $e_C = e_{1C}^W © e_{2C}^W$ and e_{1C} is of rank ρ_{1C}. By (6) and (7), $u_1 \ldots u_m$ is in VSEQ(e_{1C}) as desired.

In each case (α) and (β), (3) is true. Hence, (1) is true.

Now consider (2). Suppose that w is in the left side of (2). Then $u = \Pi_U(w)$ is in

$$\Pi_U((\mathrm{VSEQ}(\mathscr{C}_1) \cap \mathrm{VSEQ}(\mathscr{S}'_1)) \copyright (\mathrm{VSEQ}(\mathscr{C}_2) \cap \mathrm{VSEQ}(\mathscr{S}'_2)))$$
$$\subseteq \mathrm{VSEQ}(\mathscr{C}_1) \cap \mathrm{VSEQ}(\mathscr{S}'_1)$$

by Proposition 2.2(b). Similarly, v is in $\mathrm{VSEQ}(\mathscr{C}_2) \cap \mathrm{VSEQ}(\mathscr{S}'_2)$. Two cases arise.

Case (γ): $m \leq \rho(T)$. Since u is in $\mathrm{VSEQ}(\mathscr{S}'_1)$ and $|u| = m \leq \rho(T)$, u is in $\mathscr{S}'_1$. Similarly, v is in $\mathscr{S}'_2$. Hence, $w = u \copyright v$ is in

$$\mathscr{S}'_1 \copyright \mathscr{S}'_2 = \mathscr{S} \subseteq \mathrm{VSEQ}(T) \subseteq \mathrm{VSEQ}(\mathscr{C}) \cap \mathrm{VSEQ}(\mathscr{S}).$$

Case (δ): $m > \rho(T)$. Let C be in E. Three subcases arise.

Subcase (δ1): C is in $E_1 - E_2$. Let ρ_{1C} be the rank of e_{1C}. Then $e_C = e_{1C}^U$ is of rank ρ_{1C}. For each i, $\rho_{1C} + 1 \leq i \leq m$,

$$w_{i+1}(C) = u_{i+1}(C)$$

since $\Pi_U(w_i) = u_i$

$$= e_{1C}(u_{i-\rho_{1C}}, \ldots, u_{i-1}, u_i[\langle U|C\rangle])$$

since u is in $\mathrm{VSEQ}(\mathscr{C}_1) \subseteq \mathrm{VSEQ}(e_{1C})$

$$= e_C(w_{i-\rho_{1C}}, \ldots, w_{i-1}, w_i[\langle W|C\rangle])$$

since $e_C = e_{1C}^W$. Thus, w is in $\mathrm{VSEQ}(e_C)$.

Subcase (δ2): C is in $E_2 - E_1$. By a manner similar to (δ1), w is seen to be in $\mathrm{VSEQ}(e_C)$.

Subcase (δ3): C is in $E_1 \cap E_2$. Since $w_1 \ldots w_{\rho(T)}$ is in the left side of (2), we have

(8) $\quad w_1 \ldots w_{\rho(T)}$ is in $\mathrm{VSEQ}(\mathscr{C})$

by (γ). Let ρ_{1C} and ρ_{2C} be the ranks of e_{1C} and e_{2C} respectively and $\rho_C = \max\{\rho_{1C}, \rho_{2C}\}$. Then $e_C = e_{1C}^W \copyright e_{2C}^W$ and is of rank ρ_C. Clearly, for each i, $\rho_C < \rho(T) + 1 \leq i \leq m$,

(9) $\quad w_i(C) = u_i(C) = e_{1C}(u_{i-\rho_{1C}}, \ldots, u_{i-1}, u_i[\langle U|C\rangle])$

since u is in $\mathrm{VSEQ}(\mathscr{C}_1) \subseteq \mathrm{VSEQ}(e_{1C})$ and e_{1C} is of rank ρ_{1C}

$$= e_C(w_{i-\rho_C}, \ldots, w_{i-1}, w_i[\langle W|C\rangle])$$

since $e_C = e_{1C}^W \copyright e_{2C}^W$ and is of rank ρ_C. By (8) and (9), $w_1 \ldots w_m$ is in $\mathrm{VSEQ}(e_C)$.

In each case (γ) and (δ), w is in the right side of (2). Hence, (a) holds.

Now, consider (b). The proof here is similar to, but simpler than, that of (a). Indeed, let $\mathscr{F} = \{f_A \mid A \text{ in } S_1S_2\}$, $\mathscr{F}_1 = \{f_{1A} \mid A \text{ in } S_1\}$ and $\mathscr{F}_2 = \{f_{2A} \mid A \text{ in } S_2\}$. Suppose $w = w_1 \ldots w_m$ is in $\mathrm{VSEQ}(\mathscr{F})$. For each A in S_1 and each i, $1 < i \leq m$,

$$u_i(A) = w_i(A) = f_A(w_{i-1}) = f_{1A}(\Pi_U(w_{i-1}))$$

since $f_A = f_{1A}^W$ or $f_A = f_{1A}^W \copyright f_{2A}^W$

$$= f_{1A}(u_{i-1}).$$

Thus $\boldsymbol{u}$ is in VSEQ(f_{1A}) for each A in S_1. Hence $\boldsymbol{u}$ is in VSEQ($\mathscr{F}_1$). Similarly, $\boldsymbol{v}$ is in VSEQ($\mathscr{F}_2$). Thus, $\boldsymbol{w} = \boldsymbol{u} \copyright \boldsymbol{v}$ is in VSEQ($\mathscr{F}_1$) © VSEQ($\mathscr{F}_2$).

Conversely, let $\boldsymbol{w} = \boldsymbol{u} \copyright \boldsymbol{v}$, with $\boldsymbol{u}$ in VSEQ($\mathscr{F}_1$) and $\boldsymbol{v}$ in VSEQ($\mathscr{F}_2$). Suppose A is in S_1. For each i, $1 < i \leqslant m$,

$$w_i(A) = u_i(A)$$

since $\Pi_U(w_i) = u_i$

$$= f_{1A}(u_{i-1})$$

since $\boldsymbol{u}$ is in VSEQ(f_{1A})

$$= f_A(w_{i-1})$$

since $f_A = f_{1A}^W$ or $f_A = f_{1A}^W \copyright f_{2A}^W$.

Similarly, $\boldsymbol{w}$ is in VSEQ(f_A) for each A in S_1S_2, i.e., $\boldsymbol{w}$ is in VSEQ($\mathscr{F}$). Therefore, (b) holds. □

Using the lemma, we now demonstrate the following theorem.

Theorem 3.3. *For all* CSS T_1 *and* T_2,

$$\text{VSEQ}(T_1 \copyright T_2) = \text{VSEQ}(T_1) \copyright \text{VSEQ}(T_2).$$

Proof. We shall use the notation of Lemma 3.1. We first claim that

(a) $\quad \text{VSEQ}(T_1) = \text{VSEQ}(\mathscr{E}_1) \cap \text{VSEQ}(\mathscr{F}_1) \cap \text{VSEQ}(\Sigma_1) \cap \text{VSEQ}(\mathscr{I}_1')$

and

(b) $\quad \text{VSEQ}(T_2) = \text{VSEQ}(\mathscr{E}_2) \cap \text{VSEQ}(\mathscr{F}_2) \cap \text{VSEQ}(\Sigma_2) \cap \text{VSEQ}(\mathscr{I}_2')$.

By symmetry, it suffices to establish (a).

By definition,

$$\mathscr{I}_1' = \{\boldsymbol{u} \text{ in VSEQ}(T_1) \mid |\boldsymbol{u}| \leqslant \rho(T)\}$$

and

$$\mathscr{I}_1 = \{\boldsymbol{u} \text{ in VSEQ}(T_1) \mid |\boldsymbol{u}| \leqslant \rho(T_1) \leqslant \rho(T)\}.$$

Then

$$\begin{aligned} \text{VSEQ}(\mathscr{I}_1') &= \mathscr{I}_1' \cup \{\boldsymbol{uz} \mid \boldsymbol{u} \text{ in } \mathscr{I}_1', |\boldsymbol{u}| = \rho(T)\} \\ &\subseteq \mathscr{I}_1 \cup \{\boldsymbol{uz} \mid \boldsymbol{u} \text{ in } \mathscr{I}_1, |\boldsymbol{u}| = \rho(T_1) \leqslant \rho(T)\} \\ &= \text{VSEQ}(\mathscr{I}_1). \end{aligned}$$

Hence,

$$\begin{aligned} &\text{VSEQ}(\mathscr{E}_1) \cap \text{VSEQ}(\mathscr{F}_1) \cap \text{VSEQ}(\Sigma_1) \cap \text{VSEQ}(\mathscr{I}_1') \\ &\quad \subseteq \text{VSEQ}(\mathscr{E}_1) \cap \text{VSEQ}(\mathscr{F}_1) \cap \text{VSEQ}(\Sigma_1) \cap \text{VSEQ}(\mathscr{I}_1) \\ &\quad = \text{VSEQ}(T_1). \end{aligned}$$

To see the reverse inclusion, note that

$$\mathrm{VSEQ}(T_1)=\mathscr{I}'_1\cup\{uz \text{ in } \mathrm{VSEQ}(T_1)\,|\,u \text{ in } \mathscr{I}'_1\}$$
$$\subseteq \mathrm{VSEQ}(\mathscr{I}'_1) \quad \text{by definition of } \mathrm{VSEQ}(\mathscr{I}'_1),$$

and

$$\mathrm{VSEQ}(T_1)\subseteq \mathrm{VSEQ}(\mathscr{E}_1)\cap \mathrm{VSEQ}(\mathscr{F}_1)\cap \mathrm{VSEQ}(\Sigma_1).$$

Thus,

$$\mathrm{VSEQ}(T_1)\subseteq \mathrm{VSEQ}(\mathscr{E}_1)\cap \mathrm{VSEQ}(\mathscr{F}_1)\cap \mathrm{VSEQ}(\Sigma_1)\cap \mathrm{VSEQ}(\mathscr{I}'_1)$$

and (a) holds.

Now let $T=T_1 © T_2=(\langle S_1S_2\rangle, \langle I_1I_2\rangle, \langle E_1E_2\rangle, \mathscr{E}, \mathscr{F}), \{\sigma\}, \mathscr{I})$. By the definition of cohesion of CSS, $\rho(T)=\max\{\rho(T_1), \rho(T_2)\}$. Then

$$\mathrm{VSEQ}(T_1) © \mathrm{VSEQ}(T_2)$$
$$=(\mathrm{VSEQ}(\mathscr{I}'_1)\cap \mathrm{VSEQ}(\mathscr{E}_1)\cap \mathrm{VSEQ}(\mathscr{F}_1)\cap \mathrm{VSEQ}(\Sigma_1))$$
$$© (\mathrm{VSEQ}(\mathscr{I}'_2)\cap \mathrm{VSEQ}(\mathscr{E}_2)\cap \mathrm{VSEQ}(\mathscr{F}_2)\cap \mathrm{VSEQ}(\Sigma_2))$$

by (a) and (b)

$$=((\mathrm{VSEQ}(\mathscr{I}'_1)\cap \mathrm{VSEQ}(\mathscr{E}_1)) © (\mathrm{VSEQ}(\mathscr{I}'_2)\cap \mathrm{VSEQ}(\mathscr{E}_2)))$$
$$\cap(\mathrm{VSEQ}(\mathscr{F}_1) © \mathrm{VSEQ}(\mathscr{F}_2))\cap(\mathrm{VSEQ}(\Sigma_1) © \mathrm{VSEQ}(\Sigma_2))$$

by Proposition 2.3

$$=(\mathrm{VSEQ}(\mathscr{I})\cap \mathrm{VSEQ}(\mathscr{E}))\cap \mathrm{VSEQ}(\mathscr{F})\cap \mathrm{VSEQ}(\sigma)$$

by Lemma 3.2(a)–(c)

$$=\mathrm{VSEQ}(T)=\mathrm{VSEQ}(T_1 © T_2). \qquad \square$$

Corollary. *For all* CSS $T_1,\ldots,T_n$, $\mathrm{VSEQ}(T_1 © \cdots © T_n)=©_{1\le i\le n}\mathrm{VSEQ}(T_i)$.

Suppose $T_1,\ldots,T_n$ are CSS in a distributed object-history system. The effect of the corollary to Theorem 3.3 is that the cohesion of the T_i may be viewed as a CSS describing the object history system from a centralized point of view.

We conclude the section with a comment on rank-r minimum representation and cohesion. A natural question is: if $(T_1,\ldots,T_n)$ is a rank-r minimum representation, is $(T_1 © T_2, T_3,\ldots,T_n)$ a rank-r minimum representation? The answer is no (so that a rank-r minimum representation reflects some global, rather than local, properties of a collection of CSS). Indeed, let $x=(1,2,3)$, $y=(1,4,5)$ and $z=(1,6,7)$ be in $\mathrm{SEQ}(\langle A\rangle\langle B_1\rangle\langle C_1\rangle)$, and $u=(1,8,9)$, $v=(1,10,11)$ and $w=(1,12,13)$ be in $\mathrm{SEQ}(\langle A\rangle\langle B_2\rangle\langle C_2\rangle)$. It is easy to construct (details omitted) CSS T_1, T_2 and T_3 of rank 1 over $\langle AB_1C_1\rangle$, $\langle AB_2C_2\rangle$ and $\langle AB_1B_2C_1C_2\rangle$ respectively such that $\mathrm{VSEQ}(T_1)=\{x,y,z\}$, $\mathrm{VSEQ}(T_2)=\{u,v,w\}$ and $\mathrm{VSEQ}(T_3)=\{x © u, y © v, z © w\}$. Obviously, (T_1, T_2, T_3) is a rank-1 minimum representation. Note that $\mathrm{VSEQ}(T_1 © T_2)$ consists of nine tuples and $\mathrm{VSEQ}(T_3)\subsetneq \mathrm{VSEQ}(T_1 © T_2)$. Thus, $(T_1 © T_2, T_3)$ is not a rank-1 minimum representation. ((T_3, T_3) is a rank-1 minimum representation of $(T_1 © T_2, T_3)$.)

4. Preservation of CSS properties under CSS cohesion

In the previous section, we proved that the cohesion of T_1 and T_2 has the important feature of defining the cohesion of VSEQ(T_1) and VSEQ(T_2), i.e. VSEQ(T_1 © T_2) = VSEQ(T_1) © VSEQ(T_2). In this section we shall see that if both T_1 and T_2 have certain properties, then so does T_1 © T_2. Specifically we shall show that if T_1 and T_2 are both "locally representable," respectively both "*b*-representable," then so is T_1 © T_2.

We start with the notion of "local representability", a concept introduced in [2].

Definition. A CSS $T = (\mathscr{C}, \Sigma, \mathscr{I})$ is called (k_1, k_2)-*local* if

(1) $k_1 \geqslant 2$ and $k_2 \geqslant 1$,

(2) for each u, $|u| \geqslant k_1$, u is in VSEQ($\mathscr{C}$) iff $\{w \mid |w| = k_1,\ w$ an interval of $u\} \subseteq$ VSEQ($\mathscr{C}$), and

(3) for each u, $|u| \geqslant k_2$, u is in VSEQ(Σ) iff $\{w \mid |w| = k_2,\ w$ an interval of $u\} \subseteq$ VSEQ(Σ).

T is said to be *local* if it is (k_1, k_2)-local for some k_1 and k_2. T is said to be *locally representable* if there exists a local CSS $T' = (\mathscr{C}, \Sigma', \mathscr{I})$ over $\langle U\rangle$ such that VSEQ(T') = VSEQ(T).

If $T = (\mathscr{C}, \Sigma, \mathscr{I})$ is (k_1, k_2)-local, then the maintenance of a computation-tuple sequence being in VSEQ($\mathscr{C}$) just involves checking the last k_1 computation tuples, and the maintenance of being in VSEQ(Σ) the last k_2 tuples.

If T is (k_1, k_2)-local, $k_1' \geqslant k_1$ and $k_2' \geqslant k_2$, then T is (k_1', k_2')-local.

We now turn to the problem of showing that cohesion preserves the property of being locally representable. Suppose $T_i = (((\langle S_i\rangle, \langle I_i\rangle, \langle E_i\rangle, \mathscr{C}_i, \mathscr{F}_i), \Sigma_i, \mathscr{I}_i),\ i = 1, 2$, is locally representable. If $E_1 \neq \emptyset$ and $E_2 \neq \emptyset$, then the argument is simple. We use the following result established in [1]:

(*) Let $T = (\mathscr{C}, \Sigma, \mathscr{I})$ be a CSS over $\langle S\rangle\langle I\rangle\langle E\rangle$, with $E \neq \emptyset$. Then T is locally representable iff there exist a CSS $T' = (\mathscr{C}', \emptyset, \mathscr{I}')$ over $\langle S\rangle\langle I\rangle\langle E\rangle$ such that VSEQ(T') = VSEQ(T).

By (*), there exist a CSS $T_1' = (\mathscr{C}_1', \emptyset, \mathscr{I}_1')$ over $\langle S_1 I_1 E_1\rangle$ and a $T_2' = (\mathscr{C}_2', \emptyset, \mathscr{I}_2')$ over $\langle S_2 I_2 E_2\rangle$ such that VSEQ(T_1') = VSEQ(T_1) and VSEQ(T_2') and VSEQ(T_2). Let $T = T_1'$ © T_2'. By the definition of cohesion of CSS, the constraint set of T is $\{\sigma\}$, where VSEQ(σ) = SEQ($\langle S_1 S_2 I_1 I_2 E_1 E_2\rangle$). Thus, T is local. Clearly,

$$\begin{aligned}\text{VSEQ}(T_1 © T_2) &= \text{VSEQ}(T_1) © \text{VSEQ}(T_2) \quad \text{by Theorem 3.3}\\ &= \text{VSEQ}(T_1') © \text{VSEQ}(T_2')\\ &= \text{VSEQ}(T_1' © T_2')\\ &= \text{VSEQ}(T).\end{aligned}$$

Therefore, T_1 © T_2 is locally representable.

We next establish the local representability of $T_1 © T_2$ even if one of the E_i is empty. Indeed, we derive a slightly stronger result, namely, that cohesion preserves (k_1, k_2)-localness, localness, and local-representability.

Theorem 4.1. *Let T_1 and T_2 be CSS over $\langle U\rangle$ and $\langle V\rangle$ respectively.*

(a) *If T_1 and T_2 are (k_1, k_2)-local, then so is $T_1 © T_2$.*

(b) *If T_1 and T_2 are local, then so is $T_1 © T_2$.*

(c) *If T_1 and T_2 are locally representable, then so is $T_1 © T_2$.*

Proof. It obviously suffices to just demonstrate (a). Let $T_1 = (\mathscr{C}_1, \Sigma_1, \mathscr{I}_1)$, $T_2 = (\mathscr{C}_2, \Sigma_2, \mathscr{I}_2)$ and $T = T_1 © T_2 = (\mathscr{C}, \Sigma, \mathscr{I})$. Suppose that T_1 and T_2 are (k_1, k_2)-local. To establish (a), it is enough to verify that

(1) For each w in SEQ($\langle UV\rangle$) of length at least k_1, w is in VSEQ($\mathscr{C}$) iff

$$\{y \text{ in SEQ}(\langle UV\rangle) \mid |y| = k_1, y \text{ an interval of } w\} \subseteq \text{VSEQ}(\mathscr{C}); \quad \text{and}$$

(2) For each w in SEQ($\langle UV\rangle$) of length at least k_2, w is in VSEQ(σ) iff

$$\{y \text{ in SEQ}(\langle UV\rangle) \mid |y| = k_2, y \text{ an interval of } w\} \subseteq \text{VSEQ}(\sigma).$$

We shall give the argument for (2), that for (1) being similar but more complicated notationally.

Let $w = w_1 \ldots w_m$ be in SEQ($\langle UV\rangle$), $u = u_1 \ldots u_m = \Pi_U(w)$ and $v = v_1 \ldots v_m = \Pi_V(w)$. By Proposition 2.2(c), $w = u © v$. Consider (2). Since VSEQ(σ) is interval closed, the "only-if" is obvious. Turning to the "if", suppose that

(3) $\{y \text{ in SEQ}(\langle UV\rangle) \mid |y| = k_2, y \text{ an interval of } w\} \subseteq \text{VSEQ}(\sigma)$.

It is enough to show that w is in VSEQ(σ).

Suppose u' is an interval of $u = \Pi_U(w)$ of length k_2. Then there exists an interval y of w such that $\Pi_U(y) = u'$. Let $v' = \Pi_V(y)$. Note that $|y| = |v'| = k_2$. By Proposition 2.2(c), $y = u' © v'$. By (3), y is in VSEQ(σ). Then $u' = \Pi_U(y)$ is in

$$\Pi_U(\text{VSEQ}(\sigma)) = \Pi_U(\text{VSEQ}(\Sigma_1) © \text{VSEQ}(\Sigma_2))$$

by definition of cohesion of CSS

$$\subseteq \text{VSEQ}(\Sigma_1)$$

by Proposition 2.2(b); that is,

(4) u' is in VSEQ(Σ_1).

Since u' is an arbitrary interval of u of length k_2 and T_1 is (k_1, k_2)-local, it follows from (4) that

(5) u is in VSEQ(Σ_1).

Similarly,

(6) v is in VSEQ(Σ_2).

By (5) and (6), $w = u \circledcirc v$ is in

$$\text{VSEQ}(\Sigma_1) \circledcirc \text{VSEQ}(\Sigma_2) = \text{VSEQ}(\sigma)$$

as desired. Thus, (2) holds. □

The converse to each part of Theorem 4.1 is false. Indeed, let $T_i = (\langle A\rangle, \langle B_1B_2\rangle, \langle C\rangle, \{e_C\}, \{f_A\}, \sigma_i, \mathcal{I})$, $i = 1, 2$, where

(1) the domain of each attribute is[8] $\{0, 1\}$;

(2) e_C is the function from $\text{Dom}(\langle AB_1B_2\rangle)$ to $\text{Dom}(C)$ defined for each (a, b_1, b_2) in $\text{Dom}(\langle AB_1B_2\rangle)$ by $e_C(a, b_1, b_2) = 0$ if $b_1 = b_2$ and $e_C(a, b_1, b_2) = 1$ otherwise;

(3) f_A is the function from $\text{Dom}(\langle AB_1B_2C\rangle)$ to $\text{Dom}(A)$ defined for each u in $\text{Dom}(\langle A_1B_1B_2C\rangle)$ by $f_A(u) = 1$ if $u(A) = 0$ and $f_A(u) = 0$ otherwise;

(4) for k and l in $\{1, 2\}$, $k \neq l$,

$$\begin{aligned}\text{VSEQ}(\sigma_k) = \{u_1 \ldots u_m \text{ in } \text{SEQ}(\langle AB_1B_2C\rangle) \mid\, & m \geq 1; \text{ for each } i \text{ and } j,\\ & 1 \leq i, j \leq m,\ u_i(B_l) = 0, \text{ and if } u_i(B_k) = u_j(B_k), \text{ then}\\ & u_i(C) = u_j(C)\}; \quad \text{and}\end{aligned}$$

(5) $\mathcal{I} = \{(a, 1, b, a), (a, 0, 0, 0), (a, b, 1, a) \mid a, b \text{ in } \{0, 1\}\}$.

It can be shown that $T_1 \circledcirc T_2$ is (2, 2)-local, but neither T_1 nor T_2 is locally representable.

Our second result on preservation under cohesion (of CSS) concerns "*b*-representability," a concept introduced in [1].

Notation. For u and v in $\text{SEQ}(\langle U\rangle)$, $v|u$ means that v is a subsequence of u.

Definition. For each $\mathcal{B} \subseteq \text{SEQ}(\langle U\rangle)$, let $c(\mathcal{B})$ be the constraint (over $\text{SEQ}(\langle U\rangle)$) defined by u is in $\text{VSEQ}(c(\mathcal{B}))$ if there is no y in $\mathcal{B}$ such that $y|u$. A constraint σ is called *bad subsequence* if $\sigma = c(\mathcal{B})$ for some $\mathcal{B}$. Given $k > 0$, σ is called a *k-bounded bad-subsequence constraint* if $\sigma = c(\mathcal{B})$ for some k-bounded[9] $\mathcal{B}$. A constraint is called *bounded bad-subsequence* if it is a k-bounded bad subsequence constraint for some k.

Clearly, each bad-subsequence constraint is uniform.

Definition. A CSS $T = (\mathcal{C}, \Sigma, \mathcal{I})$ is said to be *b-representable* (respectively, *k-bounded b-representable*, k some positive integer) if there exists some $T' = (\mathcal{C}, \Sigma', \mathcal{I})$ such that Σ' is a set of bad-subsequence constraints (respectively, k-bounded bad-subsequence constraints) and $\text{VSEQ}(T') = \text{VESQ}(T)$. A CSS is said to be *bounded b-representable* if it is k-bounded b-representable for some $k > 0$.

In order to establish our result on the preservation of b-representability, we need the following lemma.

[8] The example can easily be modified so that the domains are infinite.

[9] $\mathcal{B} \subseteq \text{SEQ}(\langle U\rangle)$ is *k-bounded* if $|u| \leq k$ for all u in $\mathcal{B}$.

Lemma 4.2. *Let* $\mathscr{B}_1 \subseteq \mathrm{SEQ}(\langle U\rangle)$ *and* $\mathscr{B}_2 \subseteq \mathrm{SEQ}(\langle V\rangle)$. *Then*

(a) $$\mathrm{VSEQ}(c(\mathscr{B}_1)) © \mathrm{VSEQ}(c(\mathscr{B}_2)) = \mathrm{VSEQ}(c(\mathscr{B})),$$

where $\mathscr{B} = (\mathscr{B}_1 © \mathrm{SEQ}(\langle V\rangle)) \cup (\mathrm{SEQ}(\langle U\rangle) © \mathscr{B}_2)$; *and*

(b) *If* $\mathscr{B}_1$ *and* $\mathscr{B}_2$ *are k-bounded (for some k), then*

$$\mathrm{VSEQ}(c(\mathscr{B}_1)) © \mathrm{VSEQ}(c(\mathscr{B}_2)) = \mathrm{VSEQ}(c(B^{(k)})),$$

where

$$\mathscr{B}^{(k)} = (\mathscr{B}_1 © \{v \text{ in } \mathrm{SEQ}(\langle V\rangle) \mid |v| \leq k\}) \cup (\{u \text{ in } \mathrm{SEQ}(\langle U\rangle) \mid |u| \leq k\} © \mathscr{B}_2).$$

Proof. We first note the following easily seen facts (proof omitted):

(1) If $\mathscr{S}_1 \subseteq \mathscr{S}_2$ then $\mathscr{S}_1 © \mathscr{S}_2 = \mathscr{S}_1$.
(2) $\mathrm{SEQ}(\langle UV\rangle) = \mathrm{SEQ}(\langle U\rangle) © \mathrm{SEQ}(\langle V\rangle)$.
(3) $\mathrm{VSEQ}(c(\mathscr{B}_1)) © \mathrm{SEQ}(\langle V\rangle) = \mathrm{VSEQ}(c(\mathscr{B}_1 © \mathrm{SEQ}(\langle V\rangle)))$.
(4) $\mathrm{VSEQ}(c(\mathscr{B}_2)) © \mathrm{SEQ}(\langle U\rangle) = \mathrm{VSEQ}(c(\mathscr{B}_2 © \mathrm{SEQ}(\langle U\rangle)))$.
(5) $\mathrm{VSEQ}(c(\mathscr{B}_1)) \cap \mathrm{VSEQ}(c(\mathscr{B}_2)) = \mathrm{VSEQ}(c(\mathscr{B}_1 \cup \mathscr{B}_2))$.[10]

Consider (a). Clearly,

$$\mathrm{VSEQ}(c(\mathscr{B}_1)) © \mathrm{VSEQ}(c(\mathscr{B}_2))$$
$$= [\mathrm{VSEQ}(c(\mathscr{B}_1)) © \mathrm{VSEQ}(c(\mathscr{B}_2))] © \mathrm{SEQ}(\langle UV\rangle) \quad \text{by (1)}$$
$$= [\mathrm{VSEQ}(c(\mathscr{B}_1)) © \mathrm{SEQ}(\langle UV\rangle)] © [\mathrm{SEQ}(\langle UV\rangle) © \mathrm{VSEQ}(c(\mathscr{B}_2))],$$

by idempotency, associativity and commutativity

$$= [\mathrm{VSEQ}(c(\mathscr{B}_1)) © \mathrm{SEQ}(\langle U\rangle) © \mathrm{SEQ}(\langle V\rangle)] © [\mathrm{SEQ}(\langle U\rangle) © \mathrm{SEQ}(\langle V\rangle) © \mathrm{VSEQ}(c(\mathscr{B}_2))] \quad \text{by (2)}$$
$$= [\mathrm{VSEQ}(c(\mathscr{B}_1)) © \mathrm{SEQ}(\langle V\rangle)] © [\mathrm{SEQ}(\langle U\rangle) © \mathrm{VSEQ}(c(\mathscr{B}_2))] \quad \text{by (1)}$$
$$= \mathrm{VSEQ}(c(\mathscr{B}_1 © \mathrm{SEQ}(\langle V\rangle))) © \mathrm{VSEQ}(c(\mathscr{B}_2 © \mathrm{SEQ}(\langle U\rangle))) \quad \text{by (3) and (4)}$$
$$= \mathrm{VSEQ}(c(\mathscr{B}_1 © \mathrm{SEQ}(\langle V\rangle))) \cap \mathrm{VSEQ}(c(\mathscr{B}_2 © \mathrm{SEQ}(\langle U\rangle))),$$

since both sets are over the same set of attributes

$$= \mathrm{VSEQ}(c(\mathscr{B})) \quad \text{by (5)}.$$

Hence (a) holds.

Now consider (b). Suppose $\mathscr{B}_1$ and $\mathscr{B}_2$ are k-bounded (for some k). For each set X of attributes, let

$$\mathrm{SEQ}_k(\langle X\rangle) = \{x \text{ in } \mathrm{SEQ}(\langle X\rangle) \mid |x| \leq k\}.$$

[10] This appears in Lemma 2.2 of [1] in a slightly different form.

Since, for arbitrary u and v, $u © v \neq \emptyset$ implies $|u| = |v|$, it follows that $\mathscr{B}_1 © \text{SEQ}(\langle V \rangle) = \mathscr{B}_1 © \text{SEQ}_k(\langle V \rangle)$. Similarly, $\mathscr{B}_2 © \text{SEQ}(\langle U \rangle) = \mathscr{B}_2 © \text{SEQ}_k(\langle U \rangle)$. Then

$$\begin{aligned}\mathscr{B} &= (\mathscr{B}_1 © \text{SEQ}(\langle V \rangle)) \cup (\mathscr{B}_2 © \text{SEQ}(\langle U \rangle)) \\ &= (\mathscr{B}_1 © \text{SEQ}_k(\langle V \rangle)) \cup (\mathscr{B}_2 © \text{SEQ}_k(\langle U \rangle)) \\ &= \mathscr{B}^{(k)}.\end{aligned}$$

Hence,

$$\begin{aligned}\text{VSEQ}(c(\mathscr{B}_1)) © \text{VSEQ}(c(\mathscr{B}_2)) &= \text{VSEQ}(c(\mathscr{B})) \quad \text{by (a)} \\ &= \text{VSEQ}_k(c(\mathscr{B})).\end{aligned}$$

Thus, (b) holds. □

Using the previous lemma, we now have the following theorem.

Theorem 4.3. *Let* $\mathscr{B}_1 \subseteq \text{SEQ}(\langle U \rangle)$, $\mathscr{B}_2 \subseteq \text{SEQ}(\langle V \rangle)$, $T_1 = (\mathscr{C}_1, c(\mathscr{B}_1), \mathscr{I}_1)$ *and* $T_2 = (\mathscr{C}_2, c(\mathscr{B}_2), \mathscr{I}_2)$. *Then*

(a) $T_1 © T_2$ *is of the form* $(\mathscr{C}, c(\mathscr{B}), \mathscr{I})$, *where*

$$\mathscr{B} = (\mathscr{B}_1 © \text{SEQ}(\langle V \rangle)) \cup (\text{SEQ}(\langle U \rangle) © \mathscr{B}_2); \quad \textit{and}$$

(b) *If* $\mathscr{B}_1$ *and* $\mathscr{B}_2$ *are k-bounded* (*for some* k), *then* $T_1 © T_2$ *is of the form* $(\mathscr{C}, c(\mathscr{B}^{(k)}), \mathscr{I})$, *where*

$$\mathscr{B}^{(k)} = (\mathscr{B}_1 © \{v \text{ in } \text{SEQ}(\langle V \rangle) \mid |v| \leq k\}) \cup (\{u \text{ in } \text{SEQ}(\langle U \rangle) \mid |u| \leq k\} © \mathscr{B}_2).$$

Hence, $T_1 © T_2$ *is b-representable* (*respectively k-bounded b-representable*) *if* T_1 *and* T_2 *are b-representable* (*respectively k-bounded b-representable*).

Proof. We shall use the notation in Lemma 4.2. Let $T_1 © T_2 = (\mathscr{C}, \sigma, \mathscr{I})$. Then

$$\text{VSEQ}(\sigma) = \text{VSEQ}(c(\mathscr{B}_1)) © \text{VSEQ}(c(\mathscr{B}_2))$$

by definition of cohesion

$$= \text{VSEQ}(c(\mathscr{B}))$$

by Lemma 4.2(a)

Hence, (a) holds. If $\mathscr{B}_1$ and $\mathscr{B}_2$ are k-bounded (for some k), then $\text{VSEQ}(\sigma) = \text{VSEQ}(c(\mathscr{B}^{(k)}))$ by Lemma 4.2(b). Thus (b) holds. □

The converse to Theorem 4.3 is false. Indeed, let $T_i = ((\langle A \rangle, \langle B_1 B_2 \rangle, \langle C \rangle, \{e_C\}, \{f_A\}), \sigma_i, \mathscr{I})$, $i = 1, 2$, where

(1) each domain is the set of integers;

(2) e_C is the function from $\text{Dom}(\langle AB_1B_2 \rangle)$ to $\text{Dom}(C)$ defined for each (a, b_1, b_2) in $\text{Dom}(\langle AB_1B_2 \rangle)$ by $e_C(a_1, b_1, b_2) = b_1 + b_2$;

(3) f_A is the function from Dom($\langle AB_1B_2C\rangle$) to Dom(C) defined for each u in Dom($\langle AB_1B_2C\rangle$) by $f_A(u)=u(A)+1$;

(4) For each k and l in $\{1,2\}$, $k\neq l$,

$$\mathrm{VSEQ}(\sigma_k)=\{u_1\ldots u_m \text{ in } \mathrm{SEQ}(\langle AB_1B_2C\rangle)\,|\, m\geq 1; \text{ for each } i,\ 1\leq i<m,$$
$$u_i(B_k)\neq u_{i+1}(B_k) \text{ and } u_i(B_l)<u_{i+1}(B_l)\}; \quad \text{and}$$

(5) $\mathscr{I}=\{(a,b_1,b_2,b_1+b_2)\,|\,a,b_1,b_2 \text{ integers}\}$.

It can be shown that T_1 and T_2 are not b-representable but T_1 © T_2 is of form $(\mathscr{C}, c(\mathscr{B}), \mathscr{I})$, where $\mathscr{B}$ is 2-bounded.

In connection with negative results, we note that cohesion does not preserve rank-r minimum representation, i.e. $(T_1,\ldots,T_n)$ and $(T'_1,\ldots,T'_n)$ may be rank-r minimum representations without (T_1 © $T'_1,\ldots,T_n$ © T'_n) being one. Indeed, using the CSS T_1, T_2 and T_3 mentioned at the end of Section 3, (T_1,T_2,T_3) and (T_2,T_1,T_3) are rank-1 minimum representations but (T_1 © T_2, T_2 © T_1, T_3 © T_3) is not. ((T_3,T_3,T_3) is a rank-1 minimum representation of (T_1 © T_2, T_2 © T_1, T_3 © T_3).)

In conclusion, the present paper has studied the analysis of cohesion, i.e. given $T_1,\ldots,T_n$ over $\langle U_1\rangle,\ldots,\langle U_n\rangle$ respectively, perhaps with special properties, what can be said about ©$_{1\leq i\leq n}$ VSEQ(T_i). The synthesis problem is the converse (and of importance in design), namely given T over $\langle U_1\ldots U_n\rangle$, perhaps with special properties, can one find $T_1,\ldots,T_n$ over $\langle U_1\rangle,\ldots,\langle U_n\rangle$ respectively such that ©$_{1\leq i\leq n}$ VSEQ(T_i) = VSEQ(T). We leave this problem for a future investigation.

References

[1] S. Ginsburg and M. Gyssens, Object histories which avoid certain subsequences, *Inform. and Comput.* **73** (1987) 174–206.

[2] S. Ginsburg and K. Tanaka, Computation-tuple sequences and object histories, *ACM Trans. Databases* **11** (1986) 186–212.

[3] S. Ginsburg and K. Tanaka, Interval queries on object histories, in: *Proc. 10th Internat. Conf. on Very Large Data Bases*, Singapore (1984) 208–217.

[4] S. Ginsburg and C. Tang, Projection of object histories, *Theoret. Comput. Sci.* **48** (1986) 297–328.