OPTIMAL INTERVAL PARTITIONING AT GIVEN POINTS

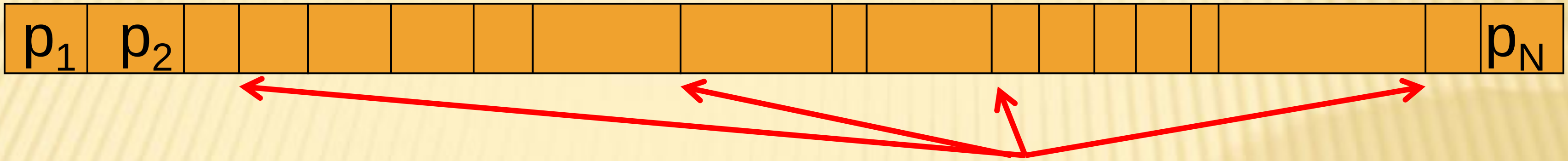
Mikhail Y. Kovalyov (Minsk, Belarus)
Sergey Sevastyanov (Novosibirsk, Russia),
Bertrand Lin (Hsinchu,Taiwan),
Feng-Jang Hwang (Sydney, Australia)

OUTLINE

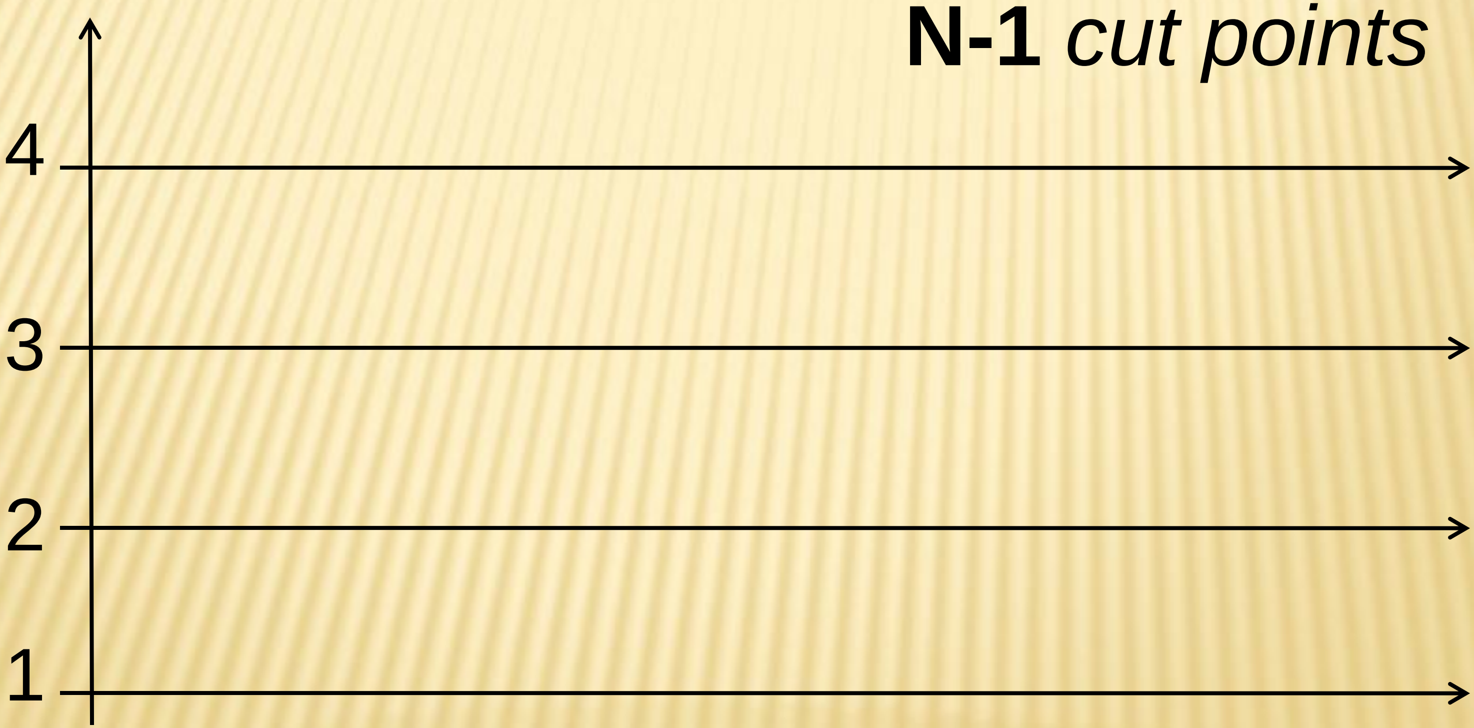
1. Formulation.
2. Scheduling interpretation.
3. Motivation.
4. Existing results.
5. Dynamic Programming 1.
6. Reducing to Series of Feasibility Problems.
7. Dynamic Programming 2 for Feasibility Problem.

1. FORMULATION

Problem **K-segments**

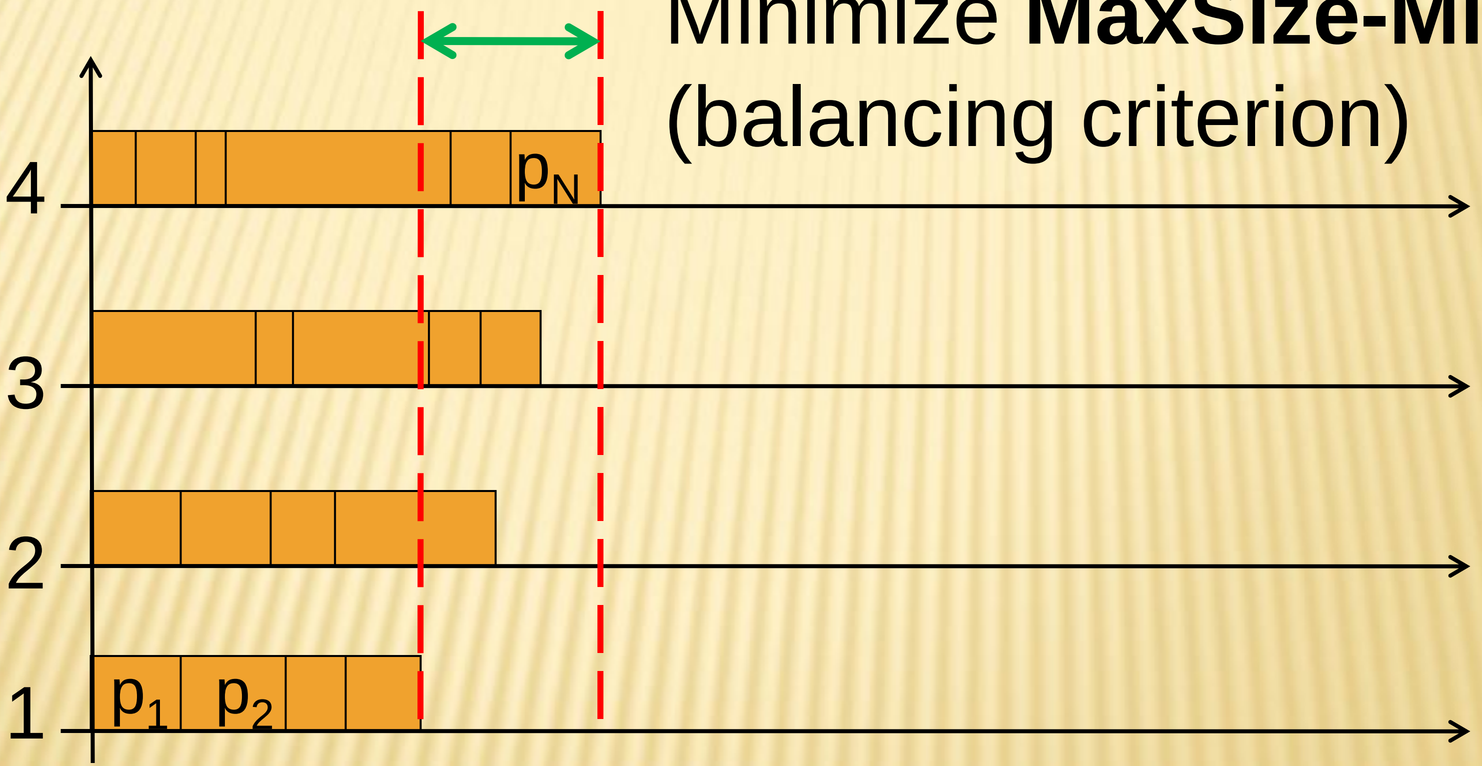


N-1 *cut points*



Cut into **K** *segments*

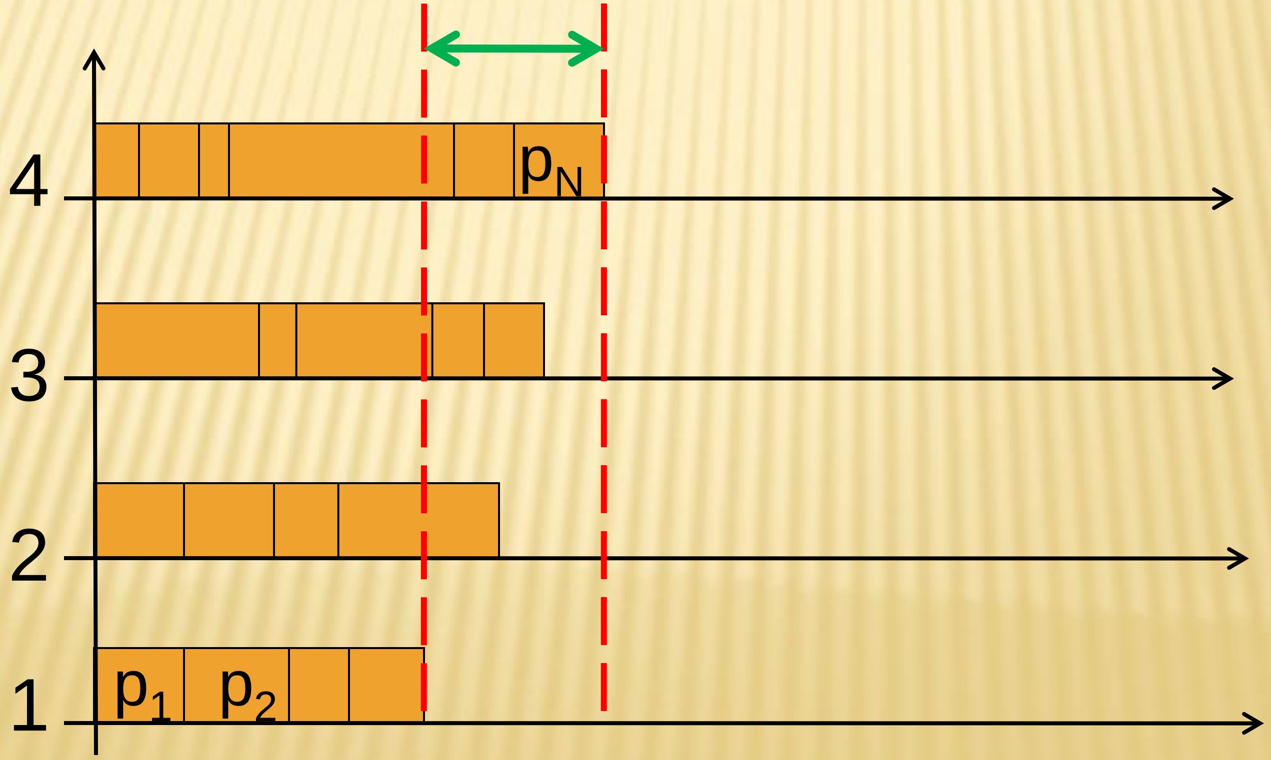
Minimize **MaxSize-MinSize**
(balancing criterion)



2. SCHEDULING INTERPRETATION

Identical Parallel Machine Scheduling:
Cut given job sequence (1,…,N) into K segments
and assign each segment to a machine

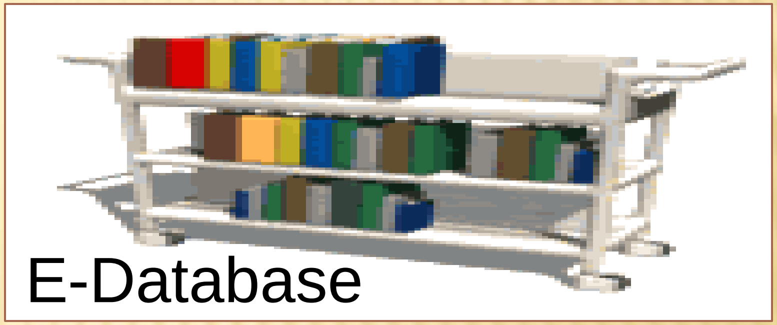
Minimize **MaxMachineLoad-MinMachineLoad**



3. MOTIVATION

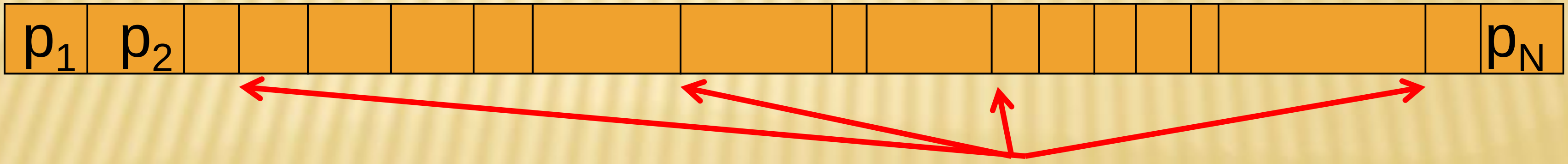
Efficient Retrieval of E-documents (Zobel et al. 1995)

Find E-docs containing (indexed) TEXT



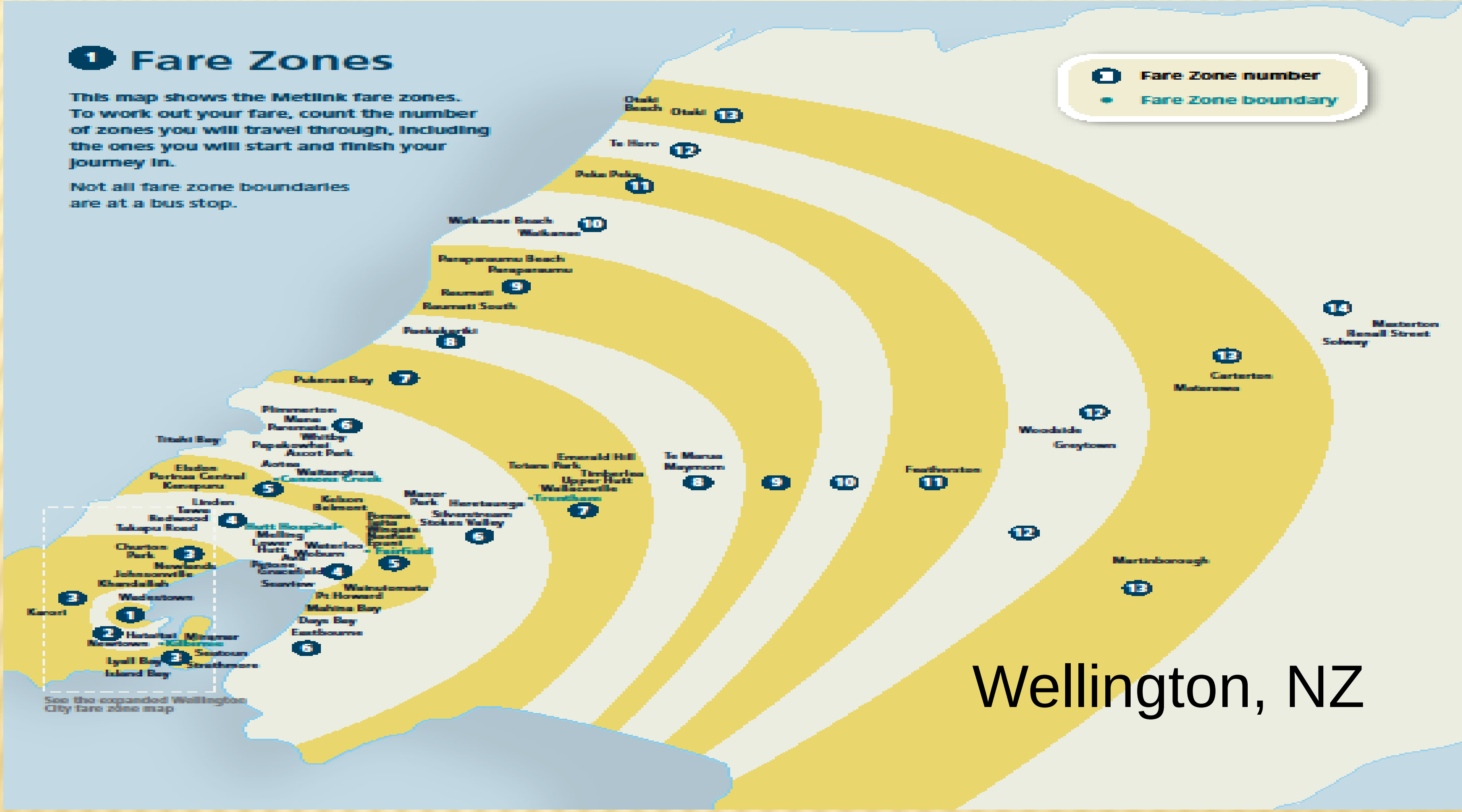
To answer queries effectively, E-docs are ranked. Ranking techniques are biased – some favor large docs, some small – and behave poorly if sizes vary. Experiments shown that partitioning large docs into **“almost equal”** parts improves retrieval effectiveness.

“War and Peace”:



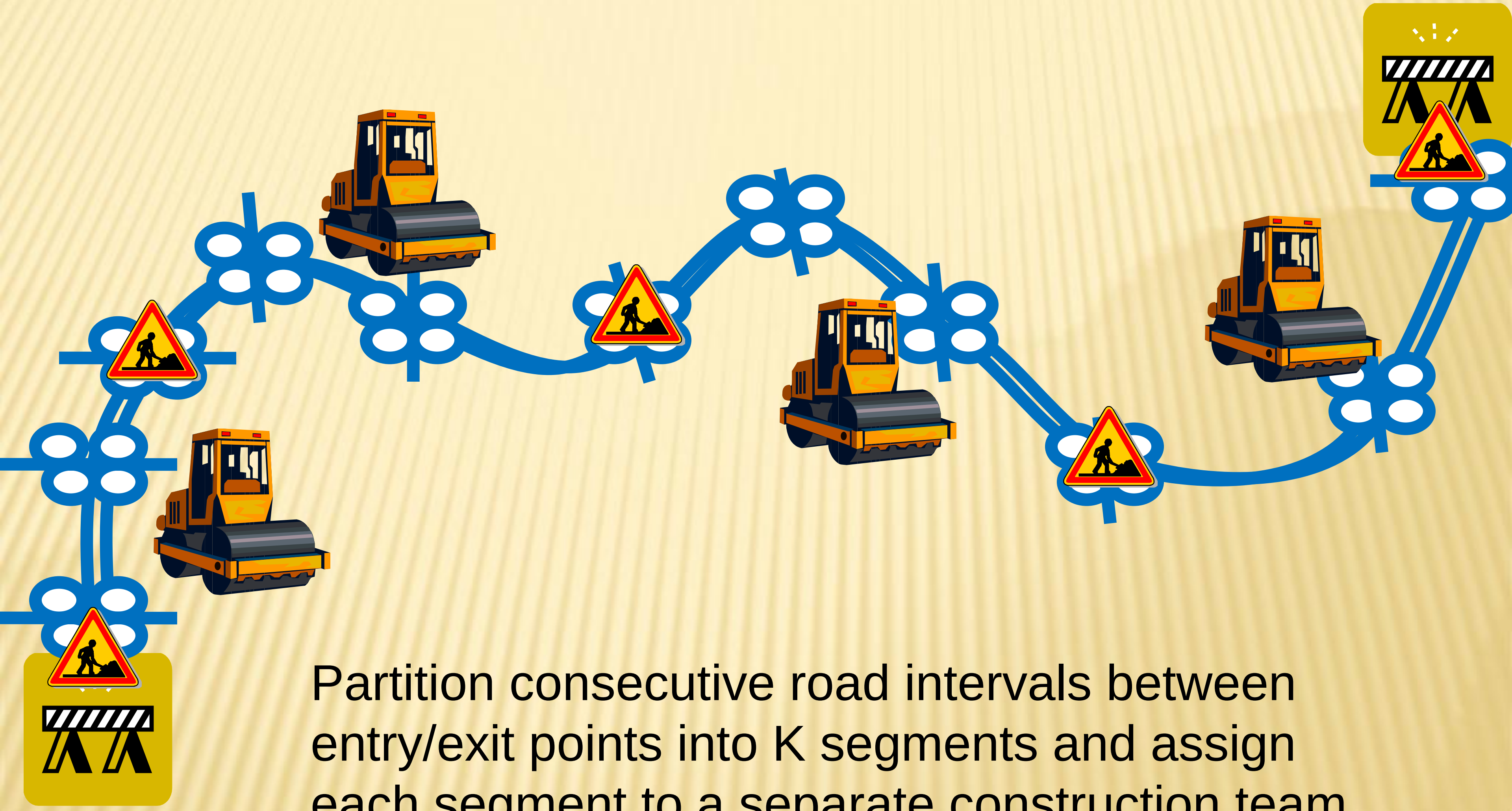
Ends of paragraphs

Fare zoning for public transport



Partition sequence of consecutive route segments into K zones so that zone lengths are maximally close to each other.

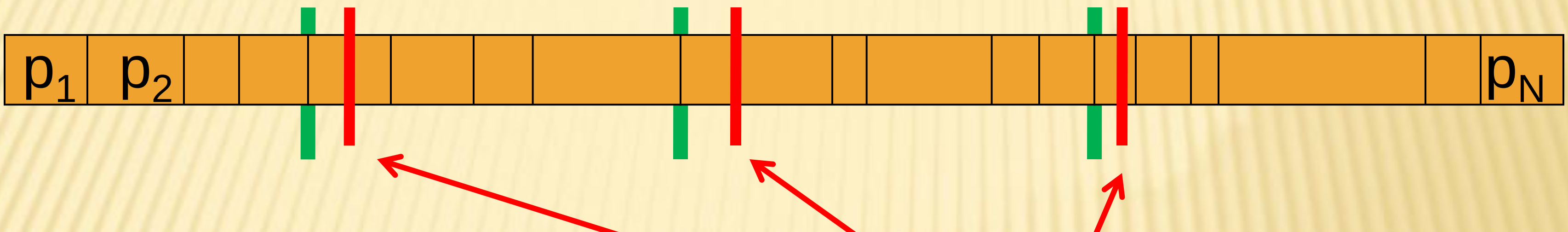
Road construction works



Partition consecutive road intervals between entry/exit points into K segments and assign each segment to a separate construction team such that team workloads are balanced.

4. EXISTING RESULTS

Imre Barany, Victor Goldberg (2013) **Block Partitions of Sequences** arXiv:1308.2452v3 [math.CO]

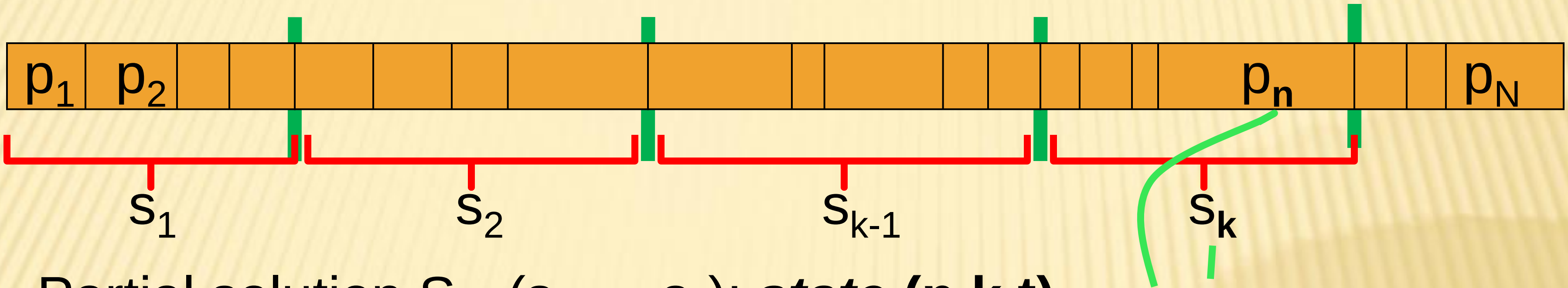


Scaled p_j≤1:

- (MaxSize-MinSize)≤2: round "ideal" cut points $k\sum_{j=1}^{N} p_j/K$, k=1,…,K-1, down to the nearest ideal cut points
- (MaxSize-MinSize)≤1: $O(KN^3)$ algorithm

- **Optimal solution: ???**

5. DYNAMIC PROGRAMMING 1



Partial solution $S_k=(s_1,\ldots,s_k)$: ***state*** **(n,k,t)**

t – future (upper bound on) max segment size,

$\Delta(n,k,t) = \min_{MaxSize \le t}(t - MinSize)$

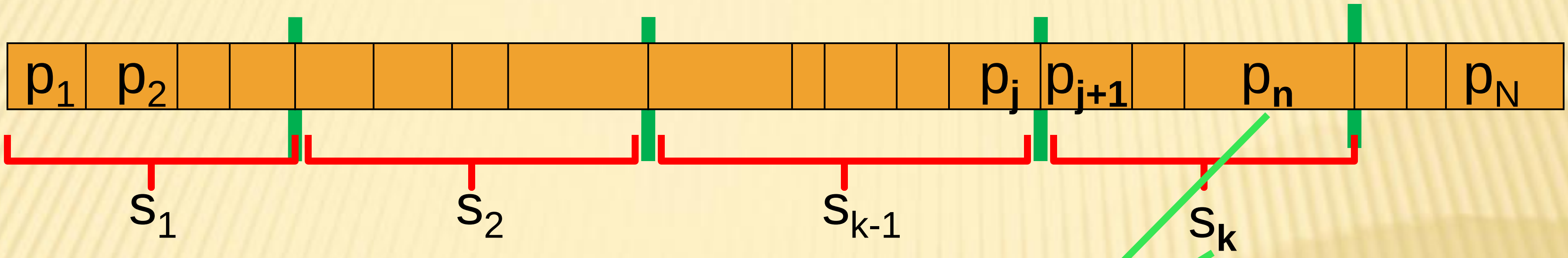
"Lack of memory" property

Richard Bellman (*Eye of the Hurricane,* autobiography): "I wanted to get across that this was dynamic, this was multi-stage, this was time-varying . . . . Let's take a word that has an absolutely precise meaning, namely dynamic, in the classical physical sense. It also has a very interesting property that it is impossible to use the word, dynamic, in a pejorative sense . . . . It was something that not even a Congressman could object to. So I used it as an umbrella for my activities."

(Courtesy the RAND Corporation)

Richard Bellman
1920–1984

Andrei A. Markov
1856–1922



Partial solution $S_k=(s_1,\ldots,s_k)$: ***state* (n,k,t)**

t – future (upper bound on) maximum segment size,

$\mathbf{\Delta(n,k,t)= \min_{MaxSize \le t}(t-MinSize)}$

$\mathbf{P(j,n)=\sum_{i=j}^{n} p_i,\ B = \{ P(j,n) \mid 1 \le j \le n \le N\},\ t \in B}$

Init: $\mathbf{\Delta(n,1,t)} = \begin{cases} \mathbf{t-P(1,n)}, & \text{if } \mathbf{t \ge P(1,n)} \\ \infty & \text{otherwise} \end{cases}$

O(KN⁴) — $O(KN^4)$

Recu: $\mathbf{\Delta(n,k,t)= \min_{\substack{0 \le j \le n, \\ P(j+1,n) \le t}} \{ \max\{t-P(j+1,n),\ \Delta(j,k-1,t)\} \}}$

$\mathbf{F^* = \min_{t \in B} \{\Delta(N,K,t)\}}$

6. REDUCING TO A SERIES OF FEASIBILITY PROBLEMS

Feasibility problem K-segments:

- Order all distinct segment sizes: $L_1 < L_2 < \ldots < L_A$, $A \le N^2$.
- For interval **$[L_a, L_b]$**, $1 \le a \le b \le A$, in **O(T)** verify
- ***Property***: **All segment sizes є $[L_a, L_b]$.** Satisfied=>**feasible**.

Solving problem **K-segments**

- Find all ***minimal*** feasible intervals $[L_a, L_b]$, $1 \le a \le b \le A$. **(# ≤ A)**
- Optimal solution corresponds to a ***minimum*** feasible interval.

O(AT)

Problem **AllMin**:

- Input: numbers 1,…,A.
- Intrl $[L_a, L_b]$, $1 \le a \le b \le A$, can satisfy ***Property,*** verified in O(T).
- Property satisfied for Intrl => satisfied for any surrounding Intrl.
- Find all minimal feasible intervals **$[L_a, L_b]$**, $1 \le a \le b \le A$.

Solving problem AllMin

				b						A
		1	2	3	4	5	6	7	8	9
	1	**0**	**0**	**1**	1	1	1	1	1	1
a	2	0	0	**1**	1	1	1	1	1	1
	3	0	0	**0**	**0**	**1**	1	1	1	1
	4	0	0	0	0	**0**	**1**	1	1	1
	5	0	0	0	0	0	**0**	**1**	1	1
	6	0	0	0	0	0	0	**1**	1	1
	7	0	0	0	0	0	0	**0**	**0**	**1**
	8	0	0	0	0	0	0	0	0	0
A	9	0	0	0	0	0	0	0	0	0

Matrix $Y=(y_{ab})$:
$y_{ab}=1$ iff $[L_a,L_b]$ feasible

Solution = red entries

Sufficient to consider **“staircase”** with ≤ 2A-1entries =>

O(AT)

A≤N² for

K-segments

7. DP2 FOR FEASIBILITY PROBLEM

Problem **Feasible(a,b)**:
All segment sizes belong to $[L_a, L_b]$?

Feasibility of segment **k** depends on last job of previous segment

Equivalent definition:

$S=(s_1,\ldots,s_K)$, $s_k=(i_{k-1}+1,\ldots,i_k)$, $k=1,\ldots,K$, $i_0=0$, is feasible iff

$P(j,n)=\sum_{i=j}^{n} p_i$

$i_k \in [A(i_{k-1}), B(i_{k-1})]$, $k=1,\ldots,K$, where

- $A(i_{k-1})=\min_{n\in[i\_(k-1)+1,N]}\{ n \mid P(i_{k-1}+1,n)\geq L_a \}$
- $B(i_{k-1})=\max_{n\in[i\_(k-1)+1,N]}\{ n \mid P(i_{k-1}+1,n)\leq L_b \}$

$[A(i_{k-1}), B(i_{k-1})]$, $k=1,\ldots,K$, belong to wider set of $[A(j), B(j)]$, $j=0,1,\ldots,N-1$, (some can be ∅) calculated in O(N) and enumerated in **DP2**

$[A(i_{k-1}), B(i_{k-1})]$, $k=1,\ldots,K$, belong to wider set of $[A(j), B(j)]$, $j=0,1,\ldots,N-1$, (some can be Ø) calculated in O(N) and enumerated in **DP2**

Non-empty $[A(j), B(j)]$ satisfy
$A(j_1) \leq A(j_2) \leq \ldots \leq A(j_h)$ and $B(j_1) \leq B(j_2) \leq \ldots \leq B(j_h)$

↓

O(N) to find
$A(j_1), A(j_2), \ldots, A(j_h)$ and $B(j_1), B(j_2), \ldots, B(j_h)$

Partial solution

$S=(1,..,\mathbf{n})=(s_1,\ldots,s_{\mathbf{k}}),\ s_r=(i_{r-1}+1,\ldots,i_r),\ r=1,\ldots,\mathbf{k}$:

***feasible state* (k,n)** if

$i_r \in [A(i_{r-1}),B(i_{r-1})],\ r=1,\ldots,k$

$\mathbf{X_k}$ – set of feasible states (k,n), n=1,…,N

$P(j,n)=\sum_{i=j}^{n} p_i$

$\mathbf{n \in [A(0),B(0)]}$

Init: $\mathbf{X_1=\{\ (1,n)\ |\ L_a \le P(1,n) \le L_b,\ n=1,\ldots,N\}}$

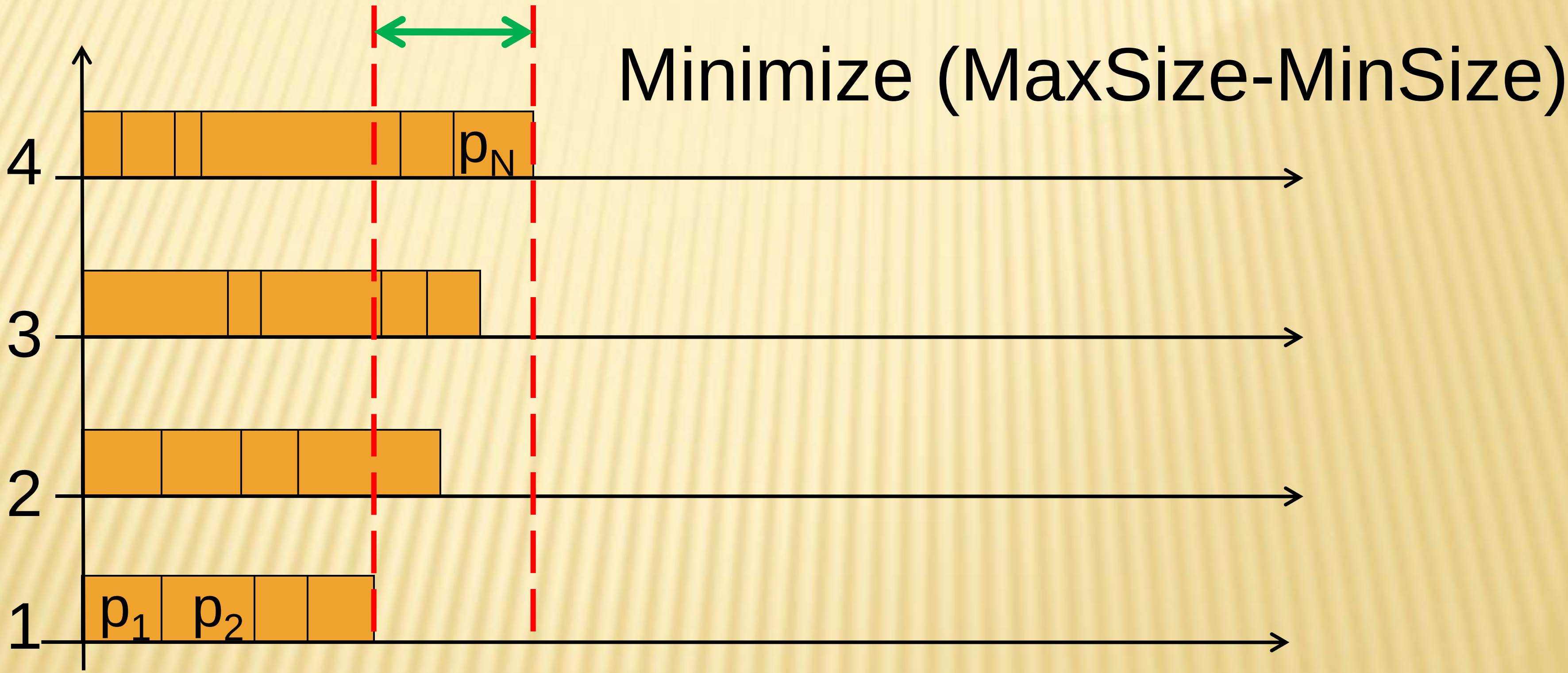
O(KN)

Recu: $\mathbf{X_k=\{\ (k,n)\ |\ n \in [A(j),B(j)],\ (k-1,j) \in X_{k-1}\ \}}$

If $\mathbf{X_K=\{(K,N)\}}$ => $\mathbf{[L_a,L_b]}$ feasible

CONCLUSION

$\leq N^2$ problems **Feasible[a,b]**, each in **O(KN)** =>
Problem **K-segments** can be solved in **O(KN^3)**



Thank you