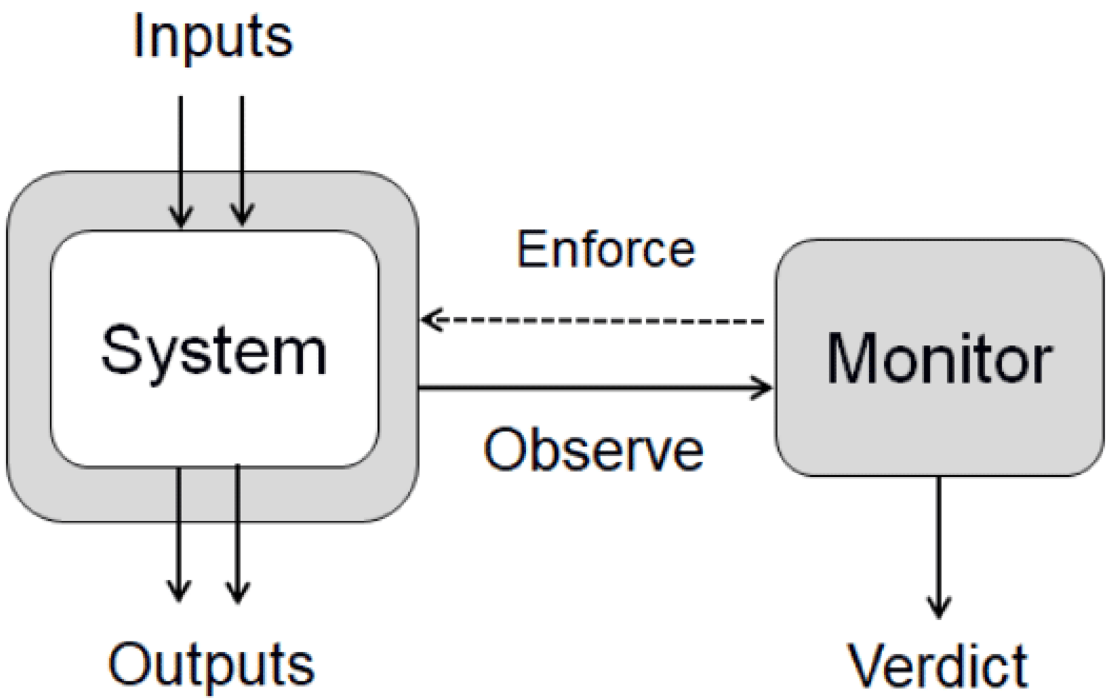
Advancing the Theory of Quantitative and Approximate Monitoring

N. Ege Saraç

Joint work with Thomas Ferreré, Thomas A. Henzinger, Nicolas Mazzocchi

Institute of Science and Technology Austria

Online Black-Box Monitoring



- Monitor runs in parallel and outputs a stream of verdicts
- Computation is deterministic and online/real-time [1]

[1] M. O. Rabin (1963). 'Real Time Computation' in *Israel J. Math. 1*.

Why Monitoring Should Become Ubiquitous

- Increasing software complexity
  - systems cannot always be model-checked or modeled
  - static verification gap?
    - better tools but harder challenges

Why Monitoring Should Become Ubiquitous

- Increasing software complexity
  - systems cannot always be model-checked or modeled
  - static verification gap?
    - better tools but harder challenges
- Increasing hardware parallelism
  - not all hardware resources will go into performance
    - some will hopefully go into assurance

Why Monitoring Should be Quantitative

- More interesting properties
  - response time, power consumption, the mode of a behavior...

Why Monitoring Should be Quantitative

- More interesting properties
  - response time, power consumption, the mode of a behavior...
- Fully universal monitoring
  - classical: boolean and existential
  - quantitative: best possible estimates at all times

Why Monitoring Should be Quantitative

- More interesting properties
  - response time, power consumption, the mode of a behavior...
- Fully universal monitoring
  - classical: boolean and existential
  - quantitative: best possible estimates at all times
- More expressive formalisms
  - monitors with integer-valued registers
  - precision-resource trade-offs

Content

1. Leveraging assumptions in monitoring [2]
2. Safety monitoring with integer-valued registers [3]
3. Limit monitoring of quantitative properties [4]
4. Abstract monitors for quantitative properties [5]

---

[2] T. A. Henzinger, N. E. Saraç (2020). 'Monitorability Under Assumptions' in *RV*.

[3] T. Ferrère, T. A. Henzinger, N. E. Saraç (2018). 'A Theory of Register Monitors' in *LICS*.

[4] T. A. Henzinger, N. E. Saraç (2021). 'Quantitative and Approximate Monitoring' in *LICS*.

[5] T. A. Henzinger, N. Mazzocchi, N. E. Saraç (2022). 'Abstract Monitors for Quantitative Specifications' in *RV*.

Monitoring Response

- Observations: $\Sigma = \{r, g, t, o\}$
- Behaviors: Σ^ω
- Property: $\Box(r \to \Diamond g)$

$$o\ t\ t\ o\ r\ t\ o\ t\ t\ t\ g\ t\ o\ o\ t\ t\ o\ t\ r\ t\ o\ t\ t\ o\ o\ o\ t\ o\ o\ t\ t\ t\ o\ o\ g\ o\ t\ o\ t\ o\ t\ o\ t\ o\ \ldots$$

- Response is
  - live (every finite behavior can be extended to satisfy the property)
  - co-live (every finite behavior can be extended to violate the property)
  - not monitorable (no finite behavior allows a definitive verdict)

Monitoring Response

- Observations: $\Sigma = \{r, g, t, o\}$
- Behaviors: Σ^ω
- Property: $\Box(r \to \Diamond g)$

$$o\ t\ t\ o\ r\ t\ o\ t\ t\ t\ g\ t\ o\ o\ t\ t\ o\ t\ r\ t\ o\ t\ t\ o\ o\ o\ t\ o\ o\ t\ t\ t\ o\ o\ g\ o\ t\ o\ t\ o\ t\ o\ t\ o\ \ldots$$

- Response is
  - live (every finite behavior can be extended to satisfy the property)
  - co-live (every finite behavior can be extended to violate the property)
  - not monitorable (no finite behavior allows a definitive verdict)

Monitoring Bounded Response

- Observations: $\Sigma = \{r, g, t, o\}$
- Behaviors: Σ^ω
- Property: $\Box(r \to \Diamond_{\leq 5}\, g)$

$$o\ t\ t\ o\ r\ t\ o\ t\ t\ t\ g\ t\ o\ o\ t\ t\ o\ t\ r\ t\ o\ t\ t\ o\ o\ o\ t\ o\ o\ t\ t\ t\ o\ o\ g\ o\ t\ o\ t\ o\ t\ o\ t\ o\ \ldots$$

- Bounded response is
  - not safe (consider the trace $r\ o\ o\ o\ o\ o\ o\ \ldots$)
  - not live (consider the finite trace $r\ t\ t\ t\ t\ t\ t$)
  - co-live (consider the extension $r\ t\ t\ t\ t\ t\ t\ \ldots$)
  - monitorable (for every finite behavior the ext. $r\ t\ t\ t\ t\ t\ t$ allows a verdict)

Monitoring Bounded Response

- Observations: $\Sigma = \{r, g, t, o\}$
- Behaviors: Σ^ω
- Property: $\Box(r \to \Diamond_{\leq 5}\, g)$

$$o\ t\ t\ o\ r\ t\ o\ t\ t\ t\ g\ t\ o\ o\ t\ t\ o\ t\ r\ t\ o\ t\ t\ o\ o\ o\ t\ o\ o\ t\ t\ t\ o\ o\ g\ o\ t\ o\ t\ o\ t\ o\ t\ o\ \ldots$$

- Bounded response is
  - not safe (consider the trace $r\ o\ o\ o\ o\ o\ o\ \ldots$)
  - not live (consider the finite trace $r\ t\ t\ t\ t\ t\ t$)
  - co-live (consider the extension $r\ t\ t\ t\ t\ t\ t\ \ldots$)
  - monitorable (for every finite behavior the ext. $r\ t\ t\ t\ t\ t\ t$ allows a verdict)

Monitoring Bounded Response

- Observations: $\Sigma = \{r, g, t, o\}$
- Assumption A: $\Box\Diamond t$
- Property: $\Box(r \rightarrow \Diamond_{\leq 5}\, g)$

$$o\ t\ t\ o\ r\ t\ o\ t\ t\ t\ g\ t\ o\ o\ t\ t\ o\ t\ r\ t\ o\ t\ t\ o\ o\ o\ t\ o\ o\ t\ t\ t\ o\ o\ g\ o\ t\ o\ t\ o\ t\ o\ t\ o\ \ldots$$

- Bounded response is
  - safe under assumption A [6]
    (every violating behavior has a finite prefix that allows a negative verdict)
  - monitorable under assumption A
    (negative verdicts can always be given in a finite number of steps)

[6] T. A. Henzinger (1992). 'Sooner is Safer than Later' in *Information Processing Letters 43(3)*.

Assumptions

- An assumption $A \subseteq \Sigma^\omega$ restricts the universe of possible behaviors
- How do assumptions arise?
  - Knowledge about the system (predictive monitoring)
  - Knowledge about the environment (e.g., $\Box\Diamond t$)
  - Information from other monitors (decentralized monitoring)
- Is there an assumption that makes a given property monitorable?

Theorem 1.

For every property P there is a co-safety assumption A_P such that P is monitorable under A_P.

Cantor Topology on Σ^ω

- Cantor metric on behaviors in Σ^ω
  - $d(f, g) = 2^{-n}$
    where n is the length of the longest common prefix of f and g
- Topological operations
  - closure = smallest closed superset
  - interior = largest open subset
  - boundary = set difference between closure and interior
- Topological characterizations
  - safe = closed; co-safe = open; live = dense [7]
  - monitorable = boundary has empty interior, i.e., nowhere dense [8]
- $A_P = \overline{\mathit{cl}(\mathit{int}(\mathit{bd}(P)))}$

---

[7] B. Alpern, F. B. Schneider (1987). 'Recognizing Safety and Liveness' in *Distributed Computing 2(3)*.

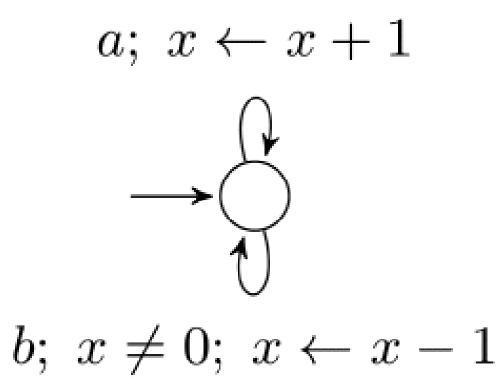
[8] V. Diekert, M. Leucker (2014). 'Topology, Monitorable Properties and Runtime Verification' in *Theor. Comput. Sci. 537*.

Safety Monitoring with Counters

- Counters: $\langle +1, -1, =0 \rangle$

$$\Sigma = \{a, b\}$$

$$P_2 = \{f \in \Sigma^\omega \mid \forall s \prec f : |s|_a \geq |s|_b\}$$



Expressiveness of Register Monitors

- $\Sigma_k = \{1, \ldots, k\}$
- $P_k = \{f \in \Sigma_k^\omega \mid \forall s \prec f : \forall i < k : |s|_{i+1} \geq |s|_i\}$

Theorem 2.

1. *For all k, the property P_{k+1} can be real-timed monitored with k counters but not with $k-1$ counters*

Expressiveness of Register Monitors

- $\Sigma_k = \{1, \dots, k\}$
- $P_k = \{f \in \Sigma_k^\omega \mid \forall s \prec f : \forall i < k : |s|_{i+1} \geq |s|_i\}$

Theorem 2.

1. *For all k, the property P_{k+1} can be real-timed monitored with k counters but not with $k-1$ counters*
2. *Any number of counters can be simulated in real-time by 4 registers with $\langle +, -, \geq \rangle$*

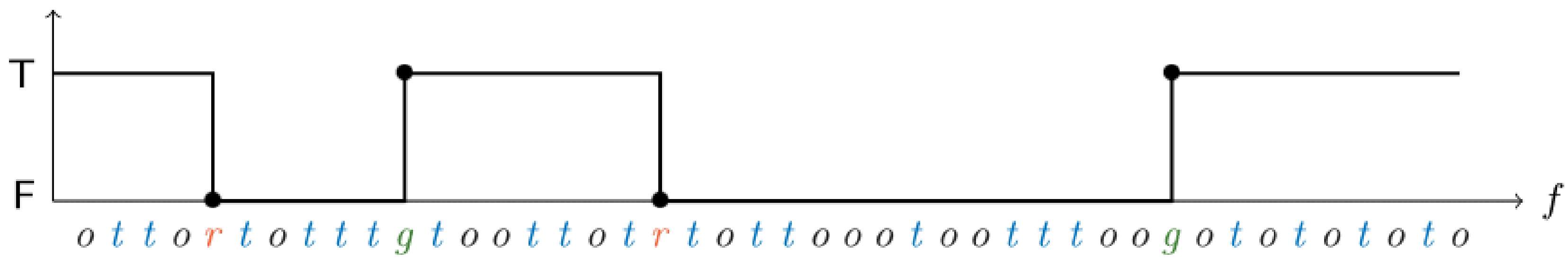
Expressiveness of Register Monitors

- $\Sigma_k = \{1, \ldots, k\}$
- $P_k = \{f \in \Sigma_k^\omega \mid \forall s \prec f : \forall i < k : |s|_{i+1} \geq |s|_i\}$
- $H = \{\#s_0\#s_1\#s_2 \ldots \mid \forall i \geq 0 : s_i \in \{0,1\}^* \wedge \forall j \neq i : s_j \neq s_i\}$
  $\hookrightarrow$ no number repeats

Theorem 2.

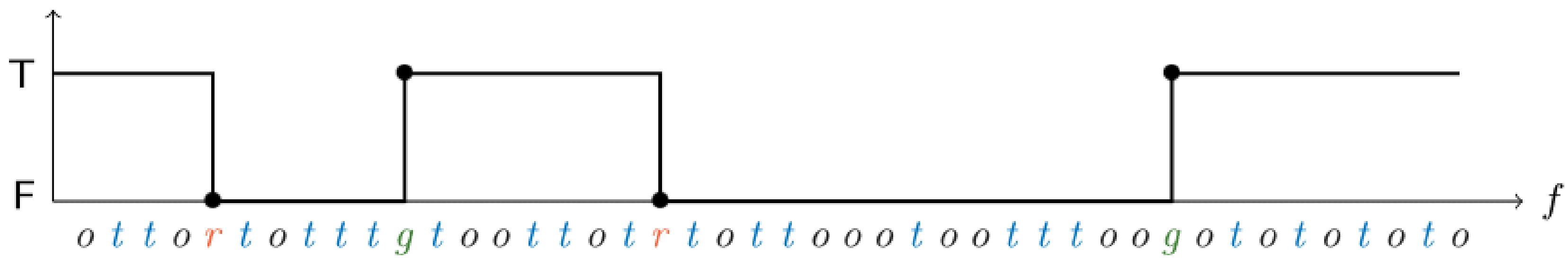
1. *For all k, the property P_{k+1} can be real-timed monitored with k counters but not with $k-1$ counters*
2. *Any number of counters can be simulated in real-time by 4 registers with $\langle +, -, \geq \rangle$*
3. *The property H cannot be real-time monitored by any number of registers with $\langle +, -, \times, \geq \rangle$*

Limit Monitoring Response



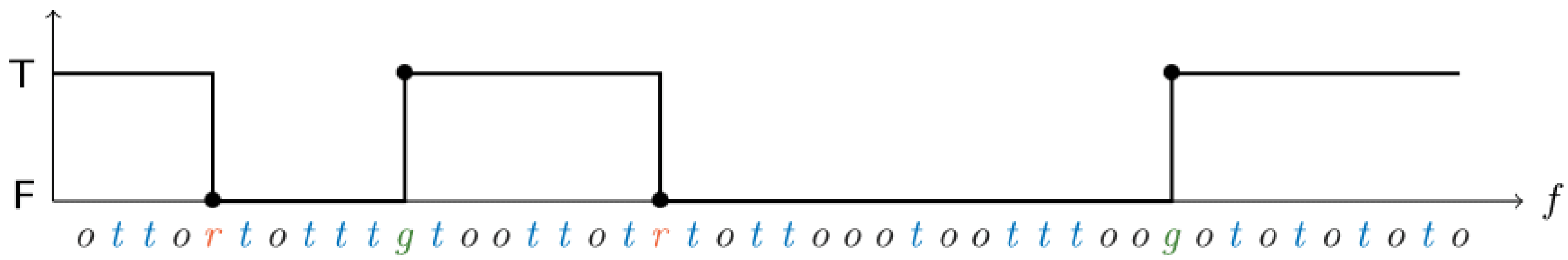
- Monitoring 'from below': $\limsup$ of verdict sequence $=$ property value
- Approximate monitoring: $\limsup$ of verdict sequence $\leq$ property value

Limit Monitoring Response



- Monitoring ‘from below’: $\limsup$ of verdict sequence $=$ property value
- Approximate monitoring: $\limsup$ of verdict sequence $\leq$ property value

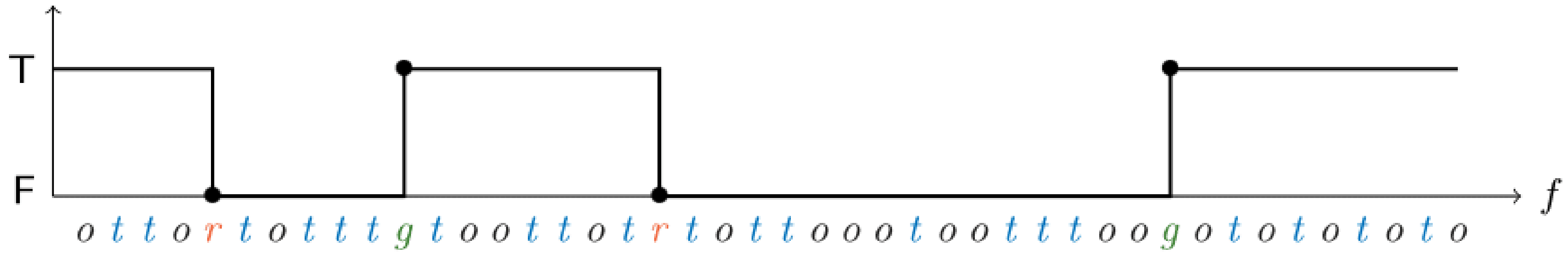
Limit Monitoring Response



- Monitoring 'from below': $\limsup$ of verdict sequence $=$ property value
- Approximate monitoring: $\limsup$ of verdict sequence $\leq$ property value

- Limit monitoring can be
  - speculative (i.e., non-monotonic)

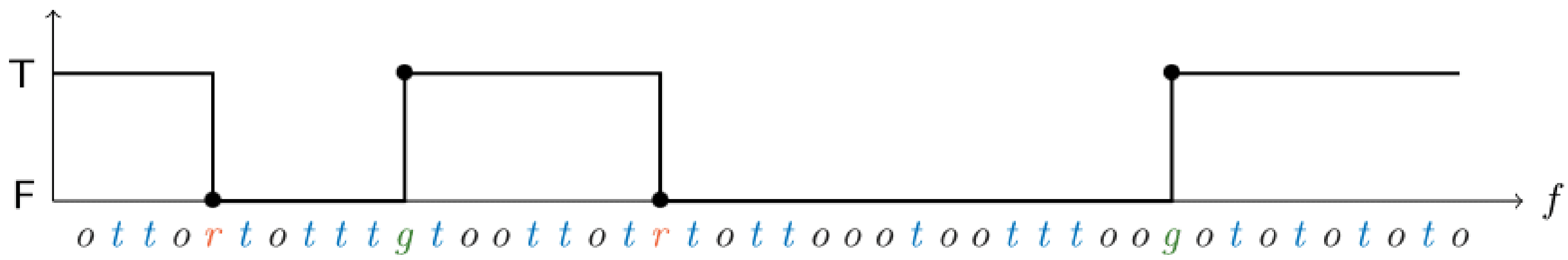
Limit Monitoring Response



- Monitoring 'from below': $\limsup$ of verdict sequence $=$ property value
- Approximate monitoring: $\limsup$ of verdict sequence $\leq$ property value

- Limit monitoring can be
  - speculative (i.e., non-monotonic)
  - universal (i.e., for all behaviors)

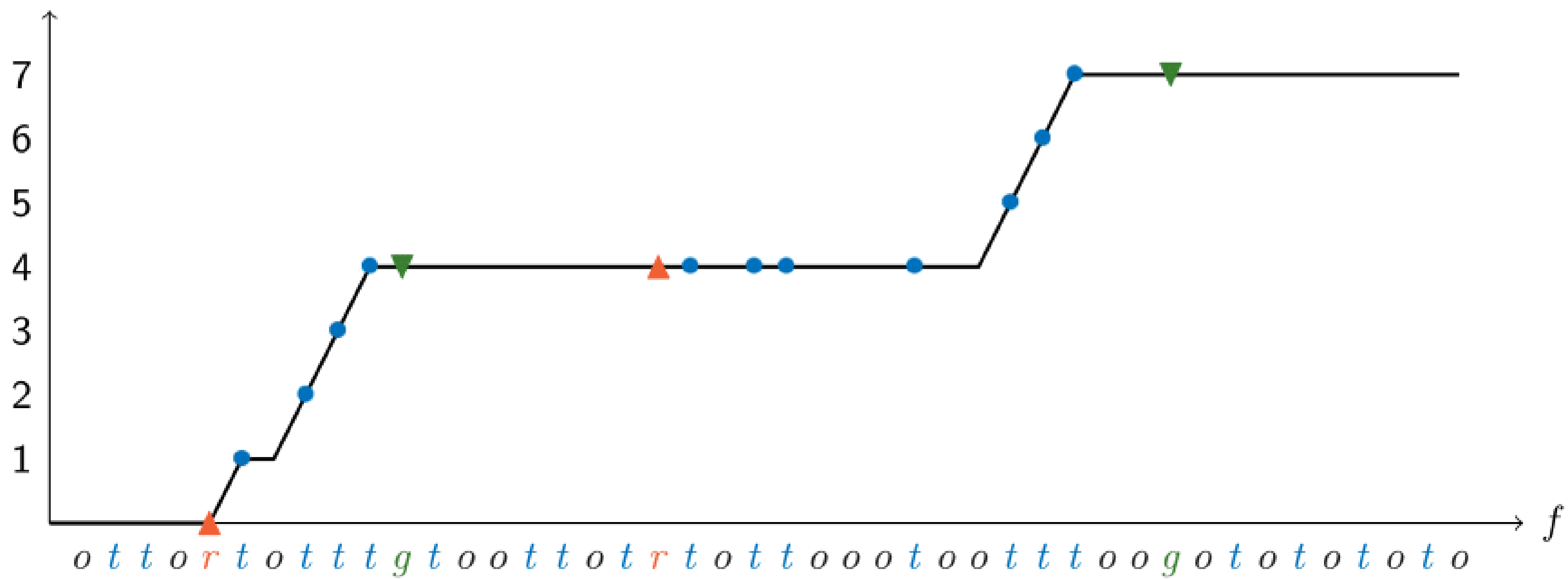
Limit Monitoring Response



- Monitoring ‘from below’: $\limsup$ of verdict sequence $=$ property value
- Approximate monitoring: $\limsup$ of verdict sequence $\leq$ property value

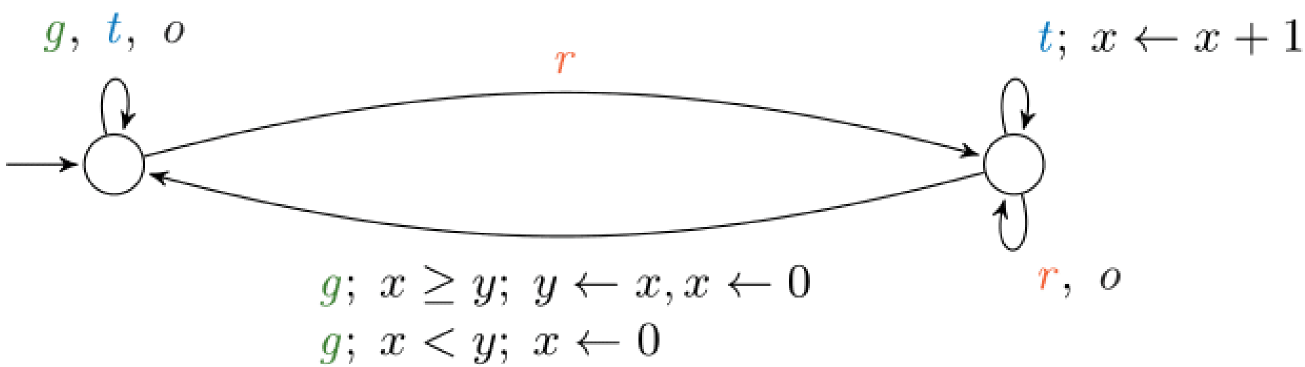
- Limit monitoring can be
  - speculative (i.e., non-monotonic)
  - universal (i.e., for all behaviors)
  - quantitative (i.e., non-boolean property values)

Monitoring Response Time



- Maximal response-time
  - monotonic
  - 2 counters

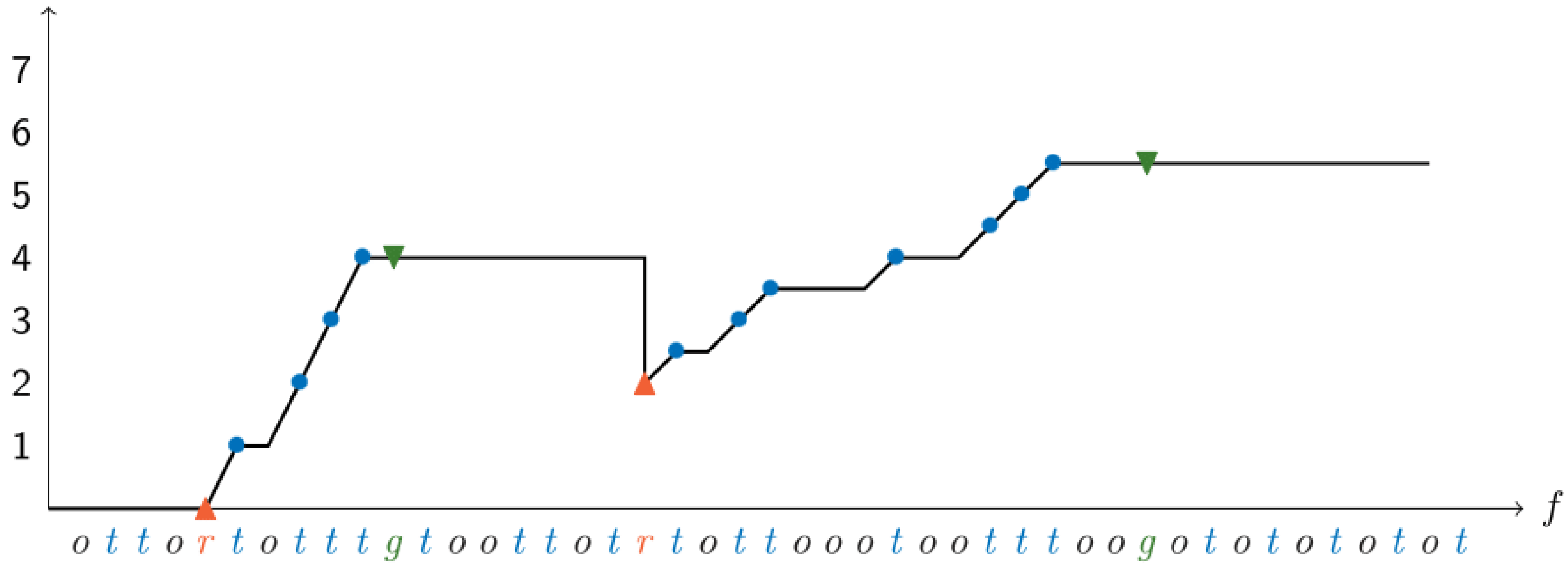
Monitoring Response Time



$$\lambda(q, \mathsf{v}) = \max(\mathsf{v}(x), \mathsf{v}(y))$$

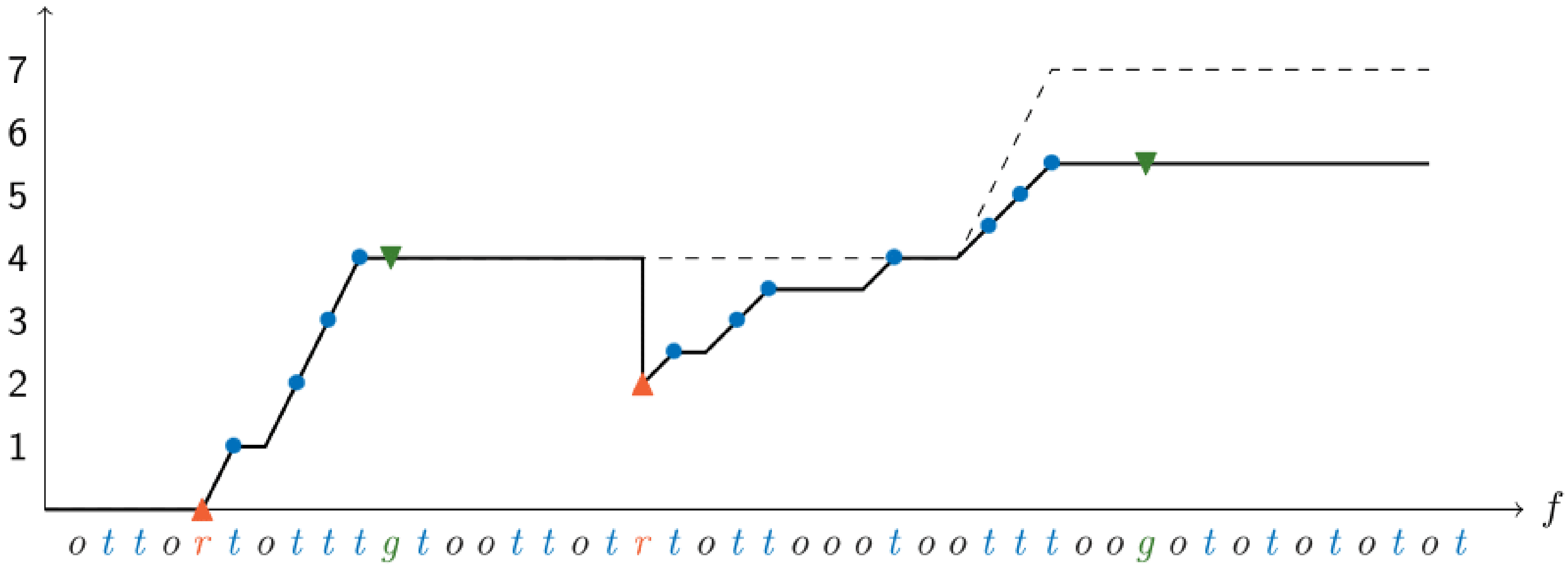
- Maximal response-time
  - monotonic
  - 2 counters

Monitoring Response Time



- Average response-time
  - non-monotonic
  - 2 registers

Monitoring Response Time



- Average response-time
  - non-monotonic
  - 2 registers

Modalities of Quantitative Monitoring

For a quantitative property $p : \Sigma^\omega \to \mathbb{D}$ and a verdict function $v : \Sigma^* \to \mathbb{D}$

- p is universally monitorable from below by v iff
  - $\forall f \in \Sigma^\omega : \limsup_{n \to \infty} v(f_{0..n}) = p(f)$
- p is existentially monitorable from below by v iff
  - $\forall f \in \Sigma^\omega : \limsup_{n \to \infty} v(f_{0..n}) \leq p(f)$
  - $\forall s \in \Sigma^* : \exists f \in \Sigma^\omega : \limsup_{n \to \infty} v(sf_{0..n}) = p(sf)$
- p is approximately monitorable from below by v iff
  - $\forall f \in \Sigma^\omega : \limsup_{n \to \infty} v(f_{0..n}) \leq p(f)$

Modalities of Quantitative Monitoring

For a quantitative property $p : \Sigma^\omega \to \mathbb{D}$ and a verdict function $v : \Sigma^* \to \mathbb{D}$

- p is universally monitorable from below by v iff
  - $\forall f \in \Sigma^\omega : \limsup_{n\to\infty} v(f_{0..n}) = p(f)$
- p is existentially monitorable from below by v iff
  - $\forall f \in \Sigma^\omega : \limsup_{n\to\infty} v(f_{0..n}) \leq p(f)$
  - $\forall s \in \Sigma^* : \exists f \in \Sigma^\omega : \limsup_{n\to\infty} v(sf_{0..n}) = p(sf)$
- p is approximately monitorable from below by v iff
  - $\forall f \in \Sigma^\omega : \limsup_{n\to\infty} v(f_{0..n}) \leq p(f)$

For verdict functions v, u that monitor p from below

- v is more precise than u iff
  - $\forall f \in \Sigma^\omega : \limsup_{n\to\infty} u(f_{0..n}) \leq \limsup_{n\to\infty} v(f_{0..n})$
  - $\exists g \in \Sigma^\omega : \limsup_{n\to\infty} u(g_{0..n}) < \limsup_{n\to\infty} v(g_{0..n})$

Modalities of Quantitative Monitoring

For a quantitative property $p : \Sigma^\omega \to \mathbb{D}$ and a verdict function $v : \Sigma^* \to \mathbb{D}$

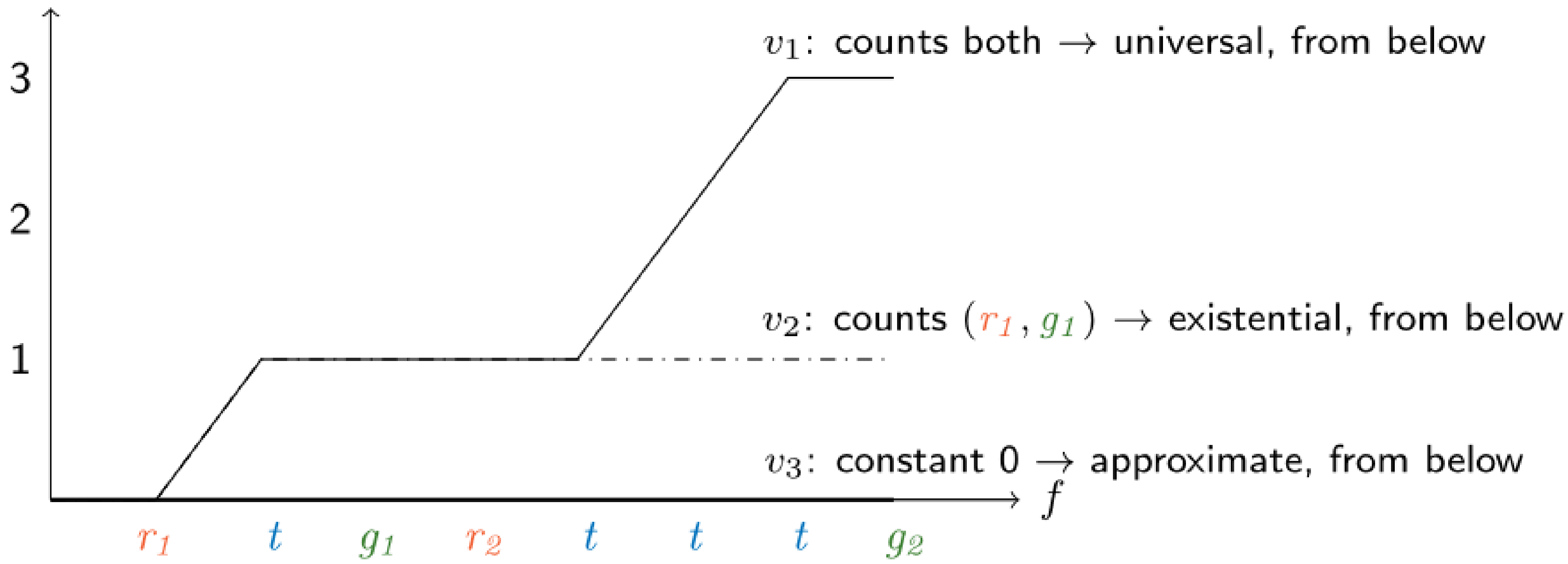
- p is universally monitorable from above by v iff
  - $\forall f \in \Sigma^\omega : \liminf_{n\to\infty} v(f_{0..n}) = p(f)$
- p is existentially monitorable from above by v iff
  - $\forall f \in \Sigma^\omega : \liminf_{n\to\infty} v(f_{0..n}) \geq p(f)$
  - $\forall s \in \Sigma^* : \exists f \in \Sigma^\omega : \liminf_{n\to\infty} v(sf_{0..n}) = p(sf)$
- p is approximately monitorable from above by v iff
  - $\forall f \in \Sigma^\omega : \liminf_{n\to\infty} v(f_{0..n}) \geq p(f)$

For verdict functions v, u that monitor p from above

- v is more precise than u iff
  - $\forall f \in \Sigma^\omega : \liminf_{n\to\infty} u(f_{0..n}) \geq \liminf_{n\to\infty} v(f_{0..n})$
  - $\exists g \in \Sigma^\omega : \liminf_{n\to\infty} u(g_{0..n}) > \liminf_{n\to\infty} v(g_{0..n})$

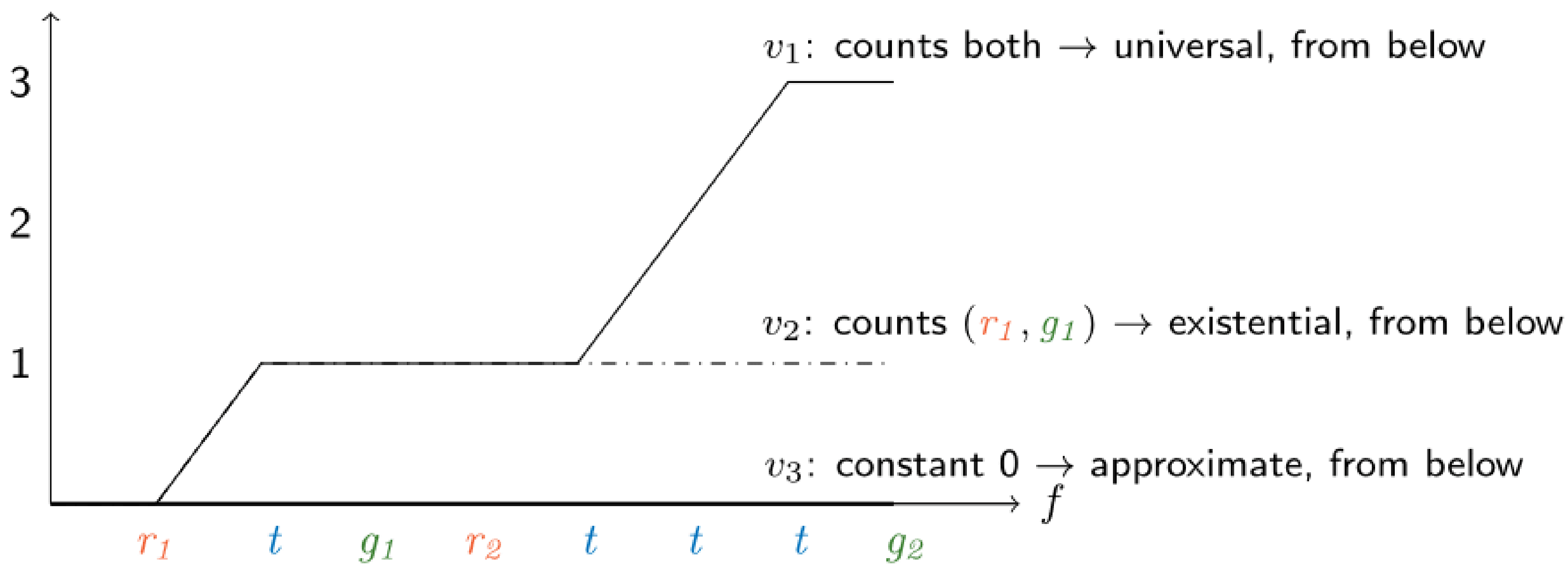
Modalities of Quantitative Monitoring

- $\Sigma = \{r_1, r_2, g_1, g_2, t, o\}$
- p_i: maximal response-time over (r_i, g_i)
- $p(f) = \max(p_1(f), p_2(f))$



Modalities of Quantitative Monitoring

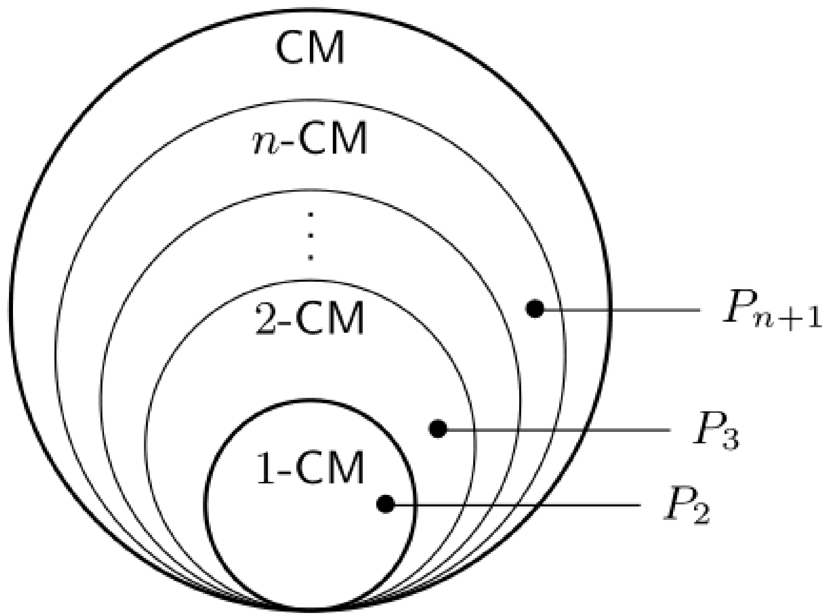
- $\Sigma = \{r_1, r_2, g_1, g_2, t, o\}$
- p_i: maximal response-time over (r_i, g_i)
- $p(f) = \max(p_1(f), p_2(f))$



- Precision: universal > existential > approximate

Approximate Counter Monitors

- Real-time counter hierarchy
  - $\Sigma_k = \{1, \ldots, k\}$
  - $P_k = \{f \in \Sigma_k^\omega \mid \forall s \prec f : \forall i < k : |s|_{i+1} \geq |s|_i\}$



Approximate Counter Monitors

- $\Sigma_k = \{1, \ldots, k\}$
- $P_k = \{f \in \Sigma_k^\omega \mid \forall s \prec f : \forall i < k : |s|_{i+1} \geq |s|_i\}$

Theorem 3.

For every k there is a quantitative property q_k such that

- q_k *requires at least k counters for universal monitoring*
- *for every $\ell < k$ each additional counter provides more precision*

Approximate Counter Monitors

- $\Sigma_k = \{1, \ldots, k\}$
- $P_k = \{f \in \Sigma_k^\omega \mid \forall s \prec f : \forall i < k : |s|_{i+1} \geq |s|_i\}$

Theorem 3.

For every k there is a quantitative property q_k such that

- *q_k requires at least k counters for universal monitoring*
- *for every $\ell < k$ each additional counter provides more precision*

Proof idea.

- Get a quantitative property q_k from P_k
  - $q_k(f) = \infty$ if $f \in P_k$
  - $q_k(f) = n$ if $f \notin P_k$ where $n :=$ len(shortest bad prefix of f)
- Approximate monitor makes "mistakes" $\rightarrow$ use the new counter to fix

$\square$

Approximate Counter Monitors

- Observations
  - The monitor needs to recognize the boolean property
  - No need to use $\Sigma_k = \{1, \ldots, k\}$ – encode using $\{0, 1, \#\}$

Theorem 4.

There is a quantitative property q such that

- q *is not universally monitorable by any counter monitor*
- *each additional counter provides more precision*

What is Missing?

- Limitations of the limit monitoring framework:
  - Restricted information from verdicts on finite prefixes
  - Fixed monitor model
  - Limited notions of resource use and precision

What is Missing?

- Limitations of the limit monitoring framework:
  - Restricted information from verdicts on finite prefixes
  - Fixed monitor model
  - Limited notions of resource use and precision

Refining Our Definitions

- Property
  - $\Phi = (\pi, \ell)$ where $\pi\colon \Sigma^* \to \mathbb{R}$ and $\ell \in \{\liminf, \limsup\}$

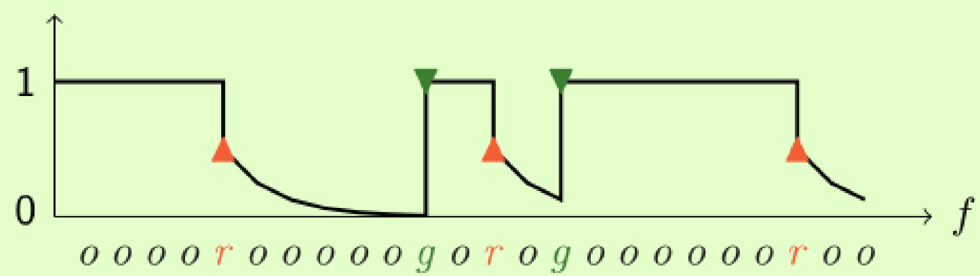
Refining Our Definitions

- Property
  - $\Phi = (\pi, \ell)$ where $\pi \colon \Sigma^* \to \mathbb{R}$ and $\ell \in \{\liminf, \limsup\}$

Example 5 (discounted response).

$\Sigma = \{r, g, o\}$, $P = \Box(r \to \Diamond g)$, $\Phi_R = (\pi, \limsup)$ where

$$\pi(s) = \begin{cases} 1 \text{ if } s \in P \\ 2^{-n} \text{ if } s \notin P \text{ where } n = |s| - |r| \text{ and } r \prec s \text{ is longest with } r \in P \end{cases}$$



Refining Our Definitions

- Abstract monitor
  - $\mathcal{M} = (\sim, \gamma)$ where $\sim$ is a right-monotonic equivalence relation and $\gamma\colon (\Sigma^*/\sim) \to \mathbb{R}$
  - $\mathcal{M}$ is a (δ, ϵ)-*monitor* for Φ iff
    - $|\Phi(s) - \mathcal{M}(s)| \leq \delta$ for all $s \in \Sigma^*$, and
    - $|\Phi(f) - \mathcal{M}(f)| \leq \epsilon$ for all $f \in \Sigma^\omega$

Refining Our Definitions

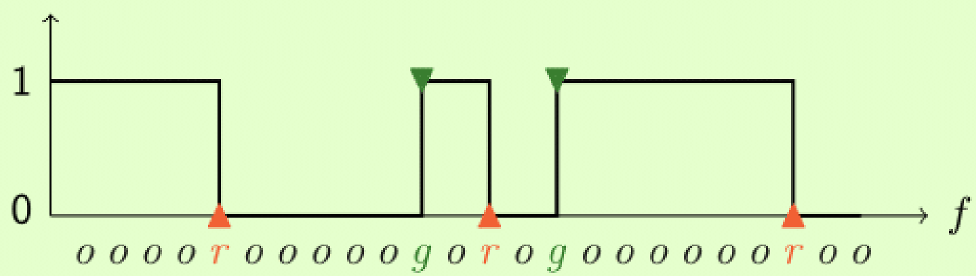
- Abstract monitor
  - $\mathcal{M} = (\sim, \gamma)$ where $\sim$ is a right-monotonic equivalence relation and $\gamma\colon (\Sigma^* / \sim) \to \mathbb{R}$
  - $\mathcal{M}$ is a (δ, ϵ)-*monitor* for Φ iff
    - $|\Phi(s) - \mathcal{M}(s)| \leq \delta$ for all $s \in \Sigma^*$, and
    - $|\Phi(f) - \mathcal{M}(f)| \leq \epsilon$ for all $f \in \Sigma^\omega$

Example 6 (discounted response).

$\Sigma = \{r, g, o\}$, $P = \Box(r \to \Diamond g)$, $\Phi_R = (\pi, \limsup)$ as above

$\mathcal{M}_B = (\sim, \gamma)$ where $s \sim r$ iff $s \in P \Leftrightarrow r \in P$,

$$\gamma([s]) = 1 \text{ if } s \in P, \text{ and } \gamma([s]) = 0 \text{ if } s \notin P$$



- $\mathcal{M}_B$ is a $(\frac{1}{2}, 0)$-monitor for Φ_R

Refining Our Definitions

- Resource use
  - Let $\mathcal{M} = (\sim, \gamma)$ and $n \in \mathbb{N}$.
    - $\mathbf{r}_n(\mathcal{M}) = |\Sigma^{\leq n}/{\sim}| - |\Sigma^{<n}/{\sim}|$

Refining Our Definitions

- Resource use
  - Let $\mathcal{M} = (\sim, \gamma)$ and $n \in \mathbb{N}$.
    - $\mathbf{r}_n(\mathcal{M}) = |\Sigma^{\leq n}/{\sim}| - |\Sigma^{<n}/{\sim}|$

Example 7 (discounted response).

$\Sigma = \{r, g, o\}$, $P = \Box(r \to \Diamond g)$, $\Phi_R = (\pi, \limsup)$
$\mathcal{M}_B$: a boolean monitor as above
$\mathcal{M}_R$: a discounting monitor that imitates Φ_R

$\mathbf{r}_0(\mathcal{M}_B) = \mathbf{r}_1(\mathcal{M}_B) = 1$ and $\mathbf{r}_n(\mathcal{M}_B) = 0$ for $n \geq 2$
$\mathbf{r}_n(\mathcal{M}_R) \geq 1$ for $n \geq 0$

$\mathcal{M}_B$ uses fewer resources than $\mathcal{M}_R$ (denoted $\mathbf{r}(\mathcal{M}_B) < \mathbf{r}(\mathcal{M}_R)$)

Refining Our Definitions

- Resource-optimality
  - Let Φ be a specification and $\mathcal{M}$ be a (δ, ϵ)-monitor for Φ.
    - $\mathcal{M}$ is *resource-optimal* for Φ iff
      $\mathcal{M}$ uses at most as many resources as any other (δ, ϵ)-monitor for Φ

Refining Our Definitions

- Resource-optimality
  - Let Φ be a specification and $\mathcal{M}$ be a (δ, ϵ)-monitor for Φ.
    - $\mathcal{M}$ is *resource-optimal* for Φ iff
      $\mathcal{M}$ uses at most as many resources as any other (δ, ϵ)-monitor for Φ
- Constructing resource-optimal monitors is difficult
  - Their structure is independent of the specification
  - Greedy approach to minimizing step-wise resource use fails

Approximate Monitoring of Maximal Response-Time

Theorem 8.

For all $\delta \in \mathbb{N}$, there exists a (δ, δ)-monitor $\mathcal{M}_\delta$ for the maximal response-time specification. Furthermore, $\mathbf{r}(\mathcal{M}_i) < \mathbf{r}(\mathcal{M}_j)$ for all $i > j$.

Proof idea.

Construct a monitor that counts response times precisely, but only stores an approximation of the maximum-so-far

- Partition $\mathbb{N}$ into intervals of size 2δ
- Make two traces equivalent if they always stay in the same interval
- Works well for traces with no pending request

□

Conclusion

Our contribution

- A framework for runtime verification which
  - supports monitors that approximate quantitative properties
  - enables comparison of such monitors w.r.t. precision and resource use

Future work

- Resource-aware decomposition of quantitative properties
- Synthesizing ‘concrete’ monitors
- Quantitative assumptions