Synthesizing Stealthy Reprogramming Attacks on Cardiac Devices

Zachary Gruber
Paul D. Schreiber High School

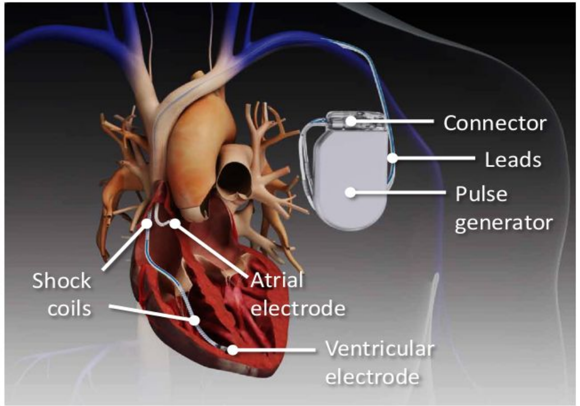
Nicola Paoletti
Stony Brook University

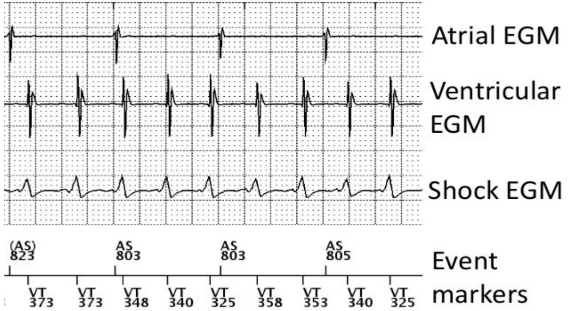
Joint work with:
Scott A Smolka, Shan Lin (Stony Brook)
Ariful Islam (CMU)
Rahul Mangharam, Houssam Abbas, Zhihao Jiang (Upenn)

CC meeting, Georgia Tech, Atlanta, 20 Apr 2018

What are ICDs?

- Implantable cardioverter defibrillator
  - 2 leads
  - 3 signals → atrial, ventricular, shock EGM
- Pacemaker and defibrillator function
- Prevent sudden death in patients
- Therapy
  - ATP - Antitachycardia pacing
  - **High-energy shocks**
- Needs to distinguish between VT and SVT
  - VT requires therapy. SVT does not.
  - **Discrimination algorithm**

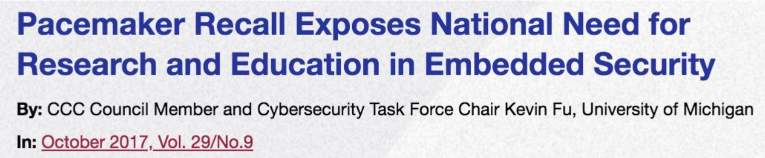




Security Concerns

- Recently security calls by the FDA

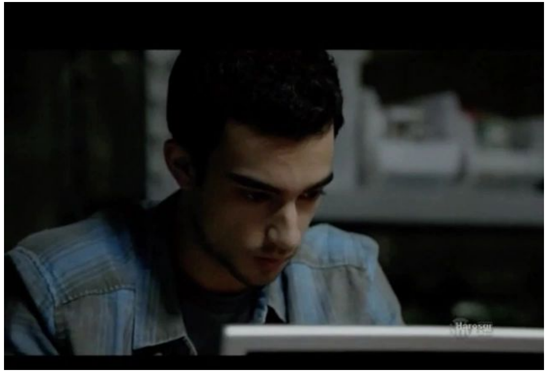
CRA
Computing Research Association



Pacemaker Recall Exposes National Need for Research and Education in Embedded Security

By: CCC Council Member and Cybersecurity Task Force Chair Kevin Fu, University of Michigan

In: October 2017, Vol. 29/No.9



Homeland, "Broken Hearts" S2E10

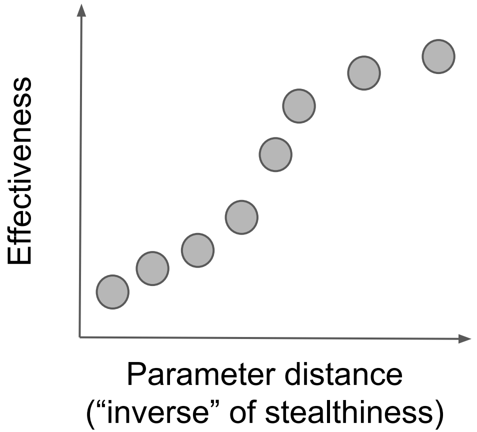
- **Study: model-based reprogramming attacks on ICDs**
  - By studying ICDs one can improve security down the road.

Related work

- Reprogramming attacks via radio (D. Halperin et al., 2008)
- Analog Spoofing (M. Reynolds et al., 2013)

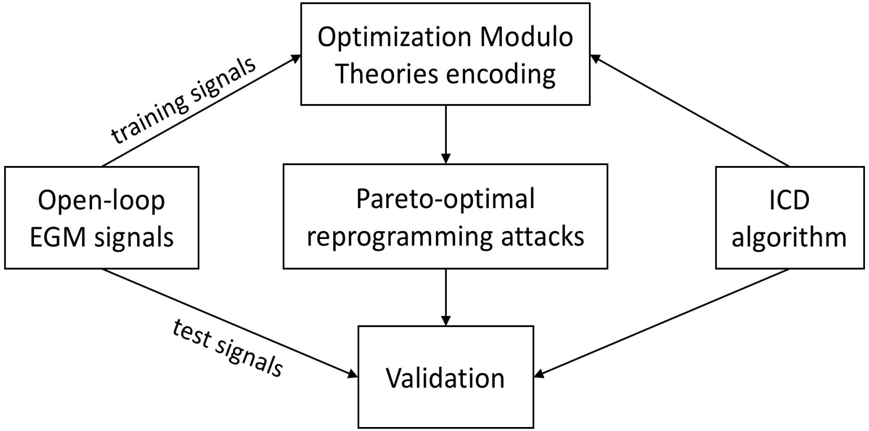
Synthesizing Stealthy Attacks on ICDs

- Reprogramming attack (manipulates ICD parameters)
- Two criteria - attack effectiveness and stealthiness
- Effectiveness:
  - Prevent necessary shocks
  - Induce unnecessary shocks
- Stealthiness:
  - Attack parameters close to the nominal parameters
  - Attack should go undetected in clinical visits → small changes mistaken by clinician’s error



Methodology Overview

- Synthesis as multi-objective optimization (stealthiness and effectiveness are contrasting objectives)
- Model of ICD discrimination algorithm
- Model-based synthetic EGM signals
    - Poor availability of real patient signals
    - Allow to tailor the attack to the victim's conditions
- Validation with unseen signals (mimicks unknown victim's EGM)


Optimization Modulo Theories encoding

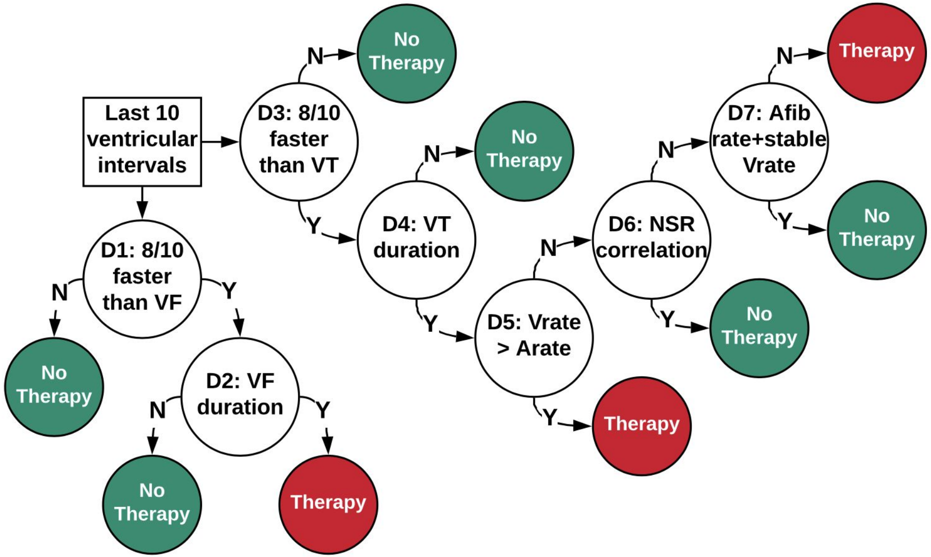
Open-loop EGM signals

Pareto-optimal reprogramming attacks

ICD algorithm

Validation

Boston Scientific ICD



Last 10 ventricular intervals

D1: 8/10 faster than VF

N → No Therapy

Y → D2: VF duration

N → No Therapy

Y → Therapy

D3: 8/10 faster than VT

N → No Therapy

Y → D4: VT duration

N → No Therapy

Y → D5: Vrate > Arate

Y → Therapy

N → D6: NSR correlation

Y → No Therapy

N → D7: Afib rate+stable Vrate

Y → No Therapy

N → Therapy

Rhythm ID discrimination algorithm

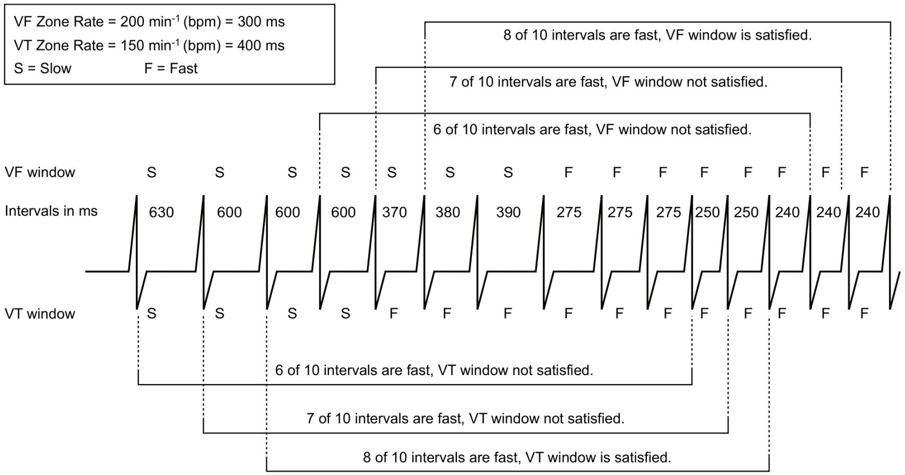
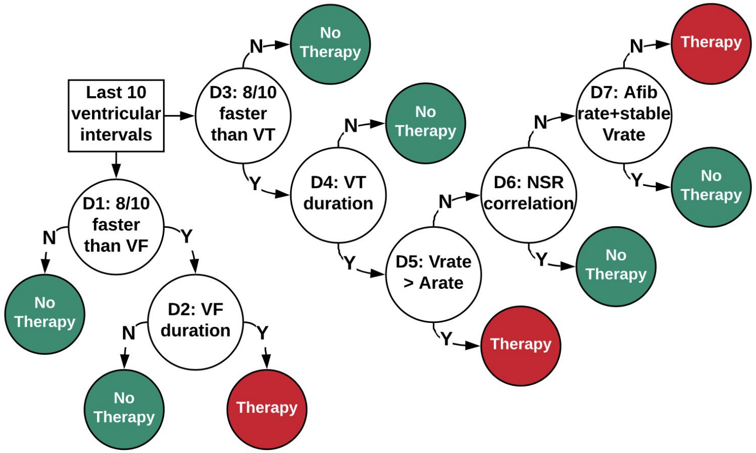

VF Zone Rate = 200 min^{-1} (bpm) = 300 ms
VT Zone Rate = 150 min^{-1} (bpm) = 400 ms
S = Slow
F = Fast
8 of 10 intervals are fast, VF window is satisfied.
7 of 10 intervals are fast, VF window not satisfied.
6 of 10 intervals are fast, VF window not satisfied.
VF window
S S S S S S S F F F F F F F F
Intervals in ms
630 600 600 600 370 380 390 275 275 275 250 250 240 240 240
VT window
S S S S F F F F F F F F F F F
6 of 10 intervals are fast, VT window not satisfied.
7 of 10 intervals are fast, VT window not satisfied.
8 of 10 intervals are fast, VT window is satisfied.

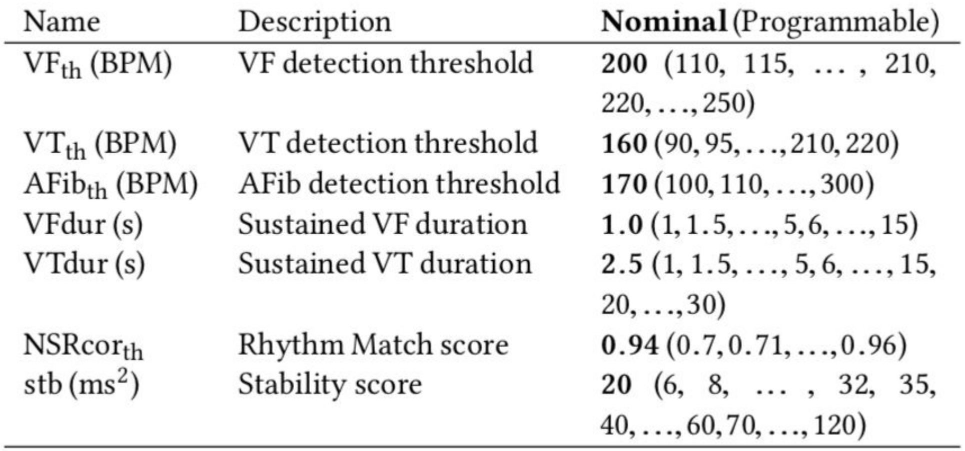
Figure 2–4. Interaction of ventricular detection windows, 2-zone configuration

Example of detection windows (BS ICD manual)

Boston Scientific ICD



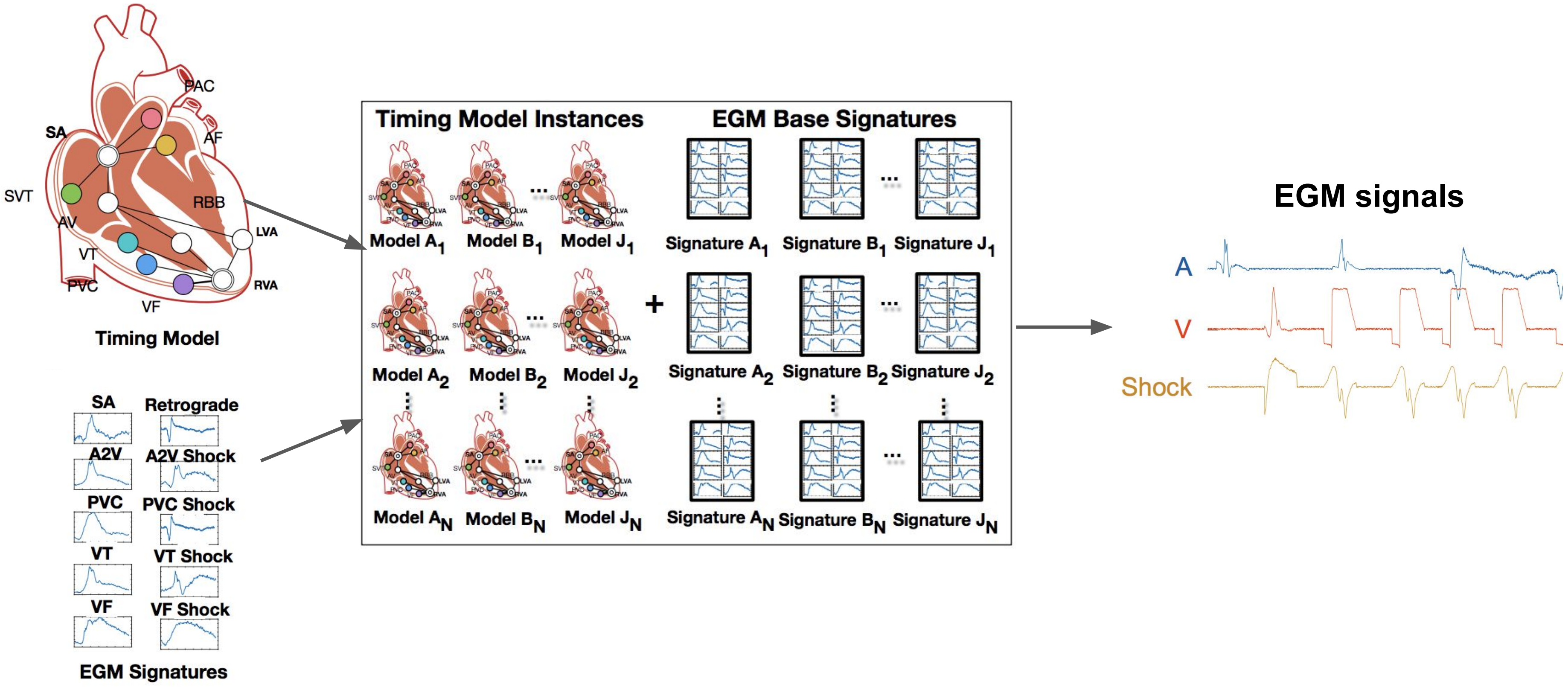
Rhythm ID discrimination algorithm



Name	Description	**Nominal** (Programmable)
VF_{th} (BPM)	VF detection threshold	**200** (110, 115, ... , 210, 220, ..., 250)
VT_{th} (BPM)	VT detection threshold	**160** (90, 95, ..., 210, 220)
$AFib_{th}$ (BPM)	AFib detection threshold	**170** (100, 110, ..., 300)
VFdur (s)	Sustained VF duration	**1.0** (1, 1.5, ..., 5, 6, ..., 15)
VTdur (s)	Sustained VT duration	**2.5** (1, 1.5, ..., 5, 6, ..., 15, 20, ..., 30)
$NSRcor_{th}$	Rhythm Match score	**0.94** (0.7, 0.71, ..., 0.96)
stb (ms^2)	Stability score	**20** (6, 8, ... , 32, 35, 40, ..., 60, 70, ..., 120)

Programmable parameters

Open-loop EGM signals (Jiang et al. EMBC 2016)



Attack effectiveness

“An attack is effective on a signal if it prevents required therapy or introduces inappropriate therapy”

$$f_e(\boxed{\mathbf{p}}, \boxed{S}) = \frac{1}{|S|} \cdot \sum_{\mathbf{s} \in S} I\left(\boxed{R_{th}(d, \mathbf{p}, \mathbf{s})} \neq \boxed{R_{th}(d, \mathbf{p}^{*}, \mathbf{s})}\right).$$

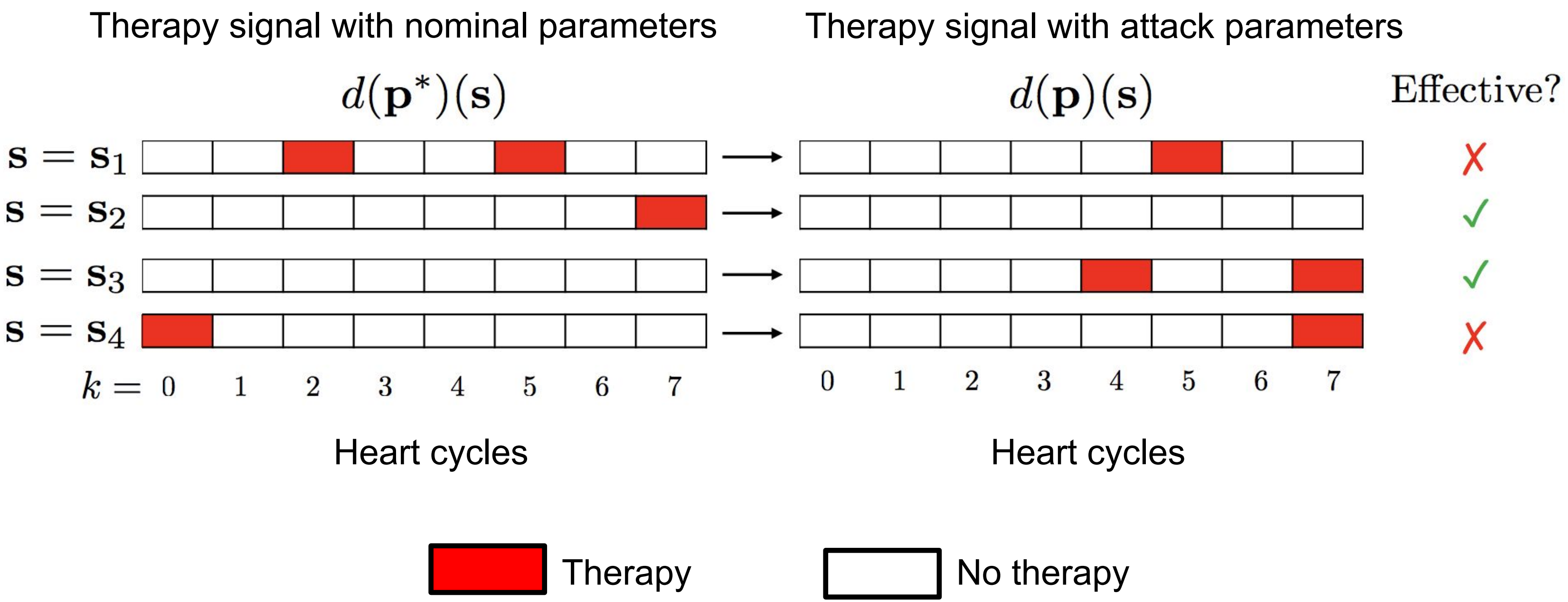
Attack parameters

Set of signals (training or test)

True iff therapy is given at any point in signal **s** under attack parameters **p**

True iff therapy is given at any point in **s** under nominal parameters **p***

Attack effectiveness (example)



	Therapy signal with nominal parameters $d(\mathbf{p}^*)(\mathbf{s})$									Therapy signal with attack parameters $d(\mathbf{p})(\mathbf{s})$								Effective?
$k =$	0	1	2	3	4	5	6	7		0	1	2	3	4	5	6	7	
$\mathbf{s} = \mathbf{s}_1$	□	□	🟥	□	□	🟥	□	□	⟶	□	□	□	□	□	🟥	□	□	✗
$\mathbf{s} = \mathbf{s}_2$	□	□	□	□	□	□	□	🟥	⟶	□	□	□	□	□	□	□	□	✓
$\mathbf{s} = \mathbf{s}_3$	□	□	□	□	□	□	□	□	⟶	□	□	□	□	🟥	□	□	🟥	✓
$\mathbf{s} = \mathbf{s}_4$	🟥	□	□	□	□	□	□	□	⟶	□	□	□	□	□	□	□	🟥	✗

Heart cycles

Heart cycles

🟥 Therapy

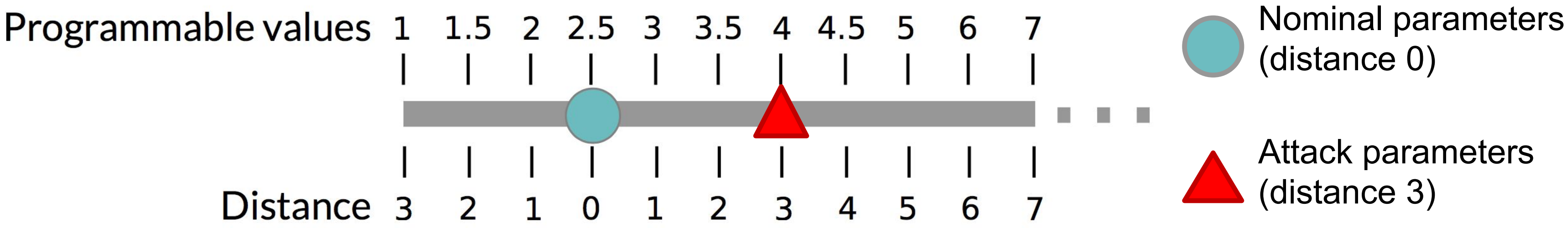
□ No therapy

Attack stealthiness

“An attack is stealthy when the deviation from the nominal parameters is small”

We quantify stealthiness as parameter distance (number of programmable values separating nominal and attack parameters)

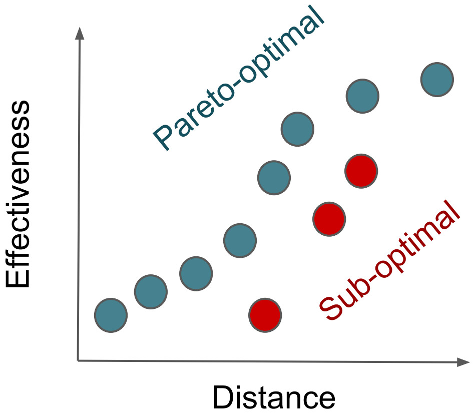
Example: parameter VT duration (s)



Synthesis of optimal stealthy attacks

Derive the set **P** of Pareto-optimal ICD parameters wrt effectiveness f_e and distance f_s objectives

$$\mathbf{P} = \{\mathbf{p} \in \mathbb{P} \mid \nexists \mathbf{p}' \in \mathbb{P}.\, (f_e(\mathbf{p}', S) > f_e(\mathbf{p}, S) \wedge f_s(\mathbf{p}') \leq f_s(\mathbf{p})) \vee (f_e(\mathbf{p}', S) \geq f_e(\mathbf{p}, S) \wedge f_s(\mathbf{p}') < f_s(\mathbf{p}))\}$$



Solution technique - optimization modulo theories (OMT)

- Optimization is challenging
  - nonlinear, non-convex, combinatorial, constrained by ICD algorithm

- SMT (SAT + theories) is well-suited to solve combinatorial problems

- **SMT encoding of BS ICD algorithm:**
  - formalization as a set FOL formulas over decidable theories (SMT QF_LIRA)
  - **Efficient encoding:** signal processing (e.g. peak detection) and nonlinear operations (e.g. correlation scores) not dependent on ICD parameters are precomputed
  - Parameter synthesis = finding a model, i.e., a SAT assignment of variables

- **OMT = SMT + precise optimization** (Bjørner et al. TACAS 2015, Sebastiani et al. CAV 2015)
  - to find the model (among all possible SAT assignments) that optimizes some objectives

OMT encoding (intuition)

BMC-like formulation:

$$\boxed{\mathrm{paramRanges}} \wedge \bigwedge_{j=1}^{|S|}\left(\boxed{\mathrm{Init}(s_{j,0})} \wedge \bigwedge_{k=0}^{N_j-1} \boxed{T(k, s_{j,k}, s_{j,k+1})}\right)$$

Constraints for programmable ranges

Initial state of ICD algorithm on j-th signal

Unrolling of transition relation describing evolution of the ICD state between heart cycles

ICD state for j-th signal and k-th heart cycle:

$$s_{j,k} \overset{\mathrm{def}}{=} (\boxed{\mathrm{VFd}_{j,k}}, \boxed{\mathrm{VTd}_{j,k}}, \boxed{\mathrm{tVF}_{j,k}}, \boxed{\mathrm{tVT}_{j,k}}) \in \mathbb{B}\times\mathbb{B}\times\mathbb{Z}^{\geq}\times\mathbb{Z}^{\geq}$$

In VF duration?

In VT duration?

Time spent in VFd

Time spent in VTd

OMT encoding (intuition)

Transition function:

$$((\neg \mathrm{VFd}_k \land \neg \mathrm{VFstart}_k) \Rightarrow \neg \mathrm{VFd}_{k+1})$$

“If outside VF duration and no VF episodes are detected, then stay outside VF duration in the next state”

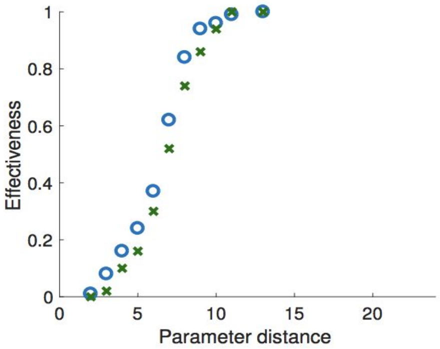
$$((\mathrm{VFstart}_k \land (\neg \mathrm{VFd}_k \lor \mathrm{VFend}_k)) \Rightarrow \mathrm{VFd}_{k+1})$$

“If a VF episode is detected and we are outside VF duration or VF duration just ended, then enter VF duration in the next state”

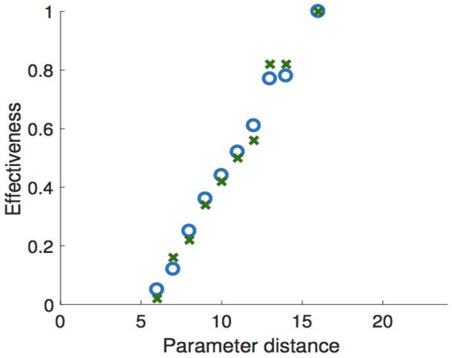
...

Evaluation, condition-specific attacks

- Use synthetic EGMs for 19 heart conditions
  - 100 EGMs for training (synthesis), 50 EGMs for validation (per condition)



Condition 10
(VT-like)



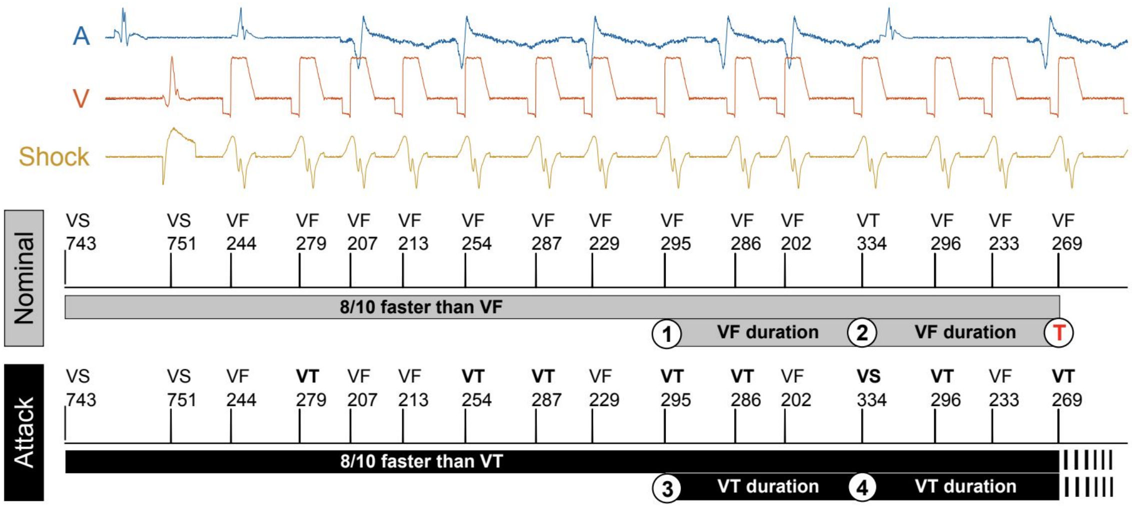
Condition 17
(VT-like)

○ Training signals ✖ Validation signals

- Attacks on VT-like conditions are all very effective
- But not all equally stealthy (see left)
- Common attack strategy:
  - Increase VT and VF detection thresholds in order to miss episodes
  - Increase VF and VT durations to reduce probability that episode is marked sustained

Evaluation, condition-specific attacks

EGM extract from condition 10 signals



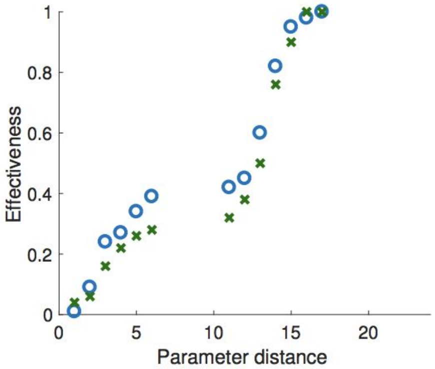
Nominal parameters:

1) VF duration start as 8/10 last ventricular intervals are below VF threshold
2) One interval is found below VF_th. Duration ends but can start right away, ending with therapy delivery (**T**)

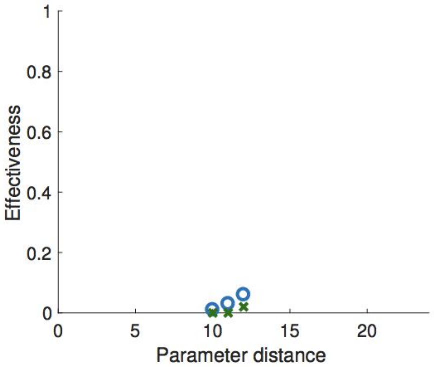
Attack parameters:

3) The episode is marked as VT and not VF (due to higher thresholds)
4) One interval is found below VT_th. VT duration ends but can start right away. Longer VT duration prevents therapy

Evaluation, condition-specific attacks



Condition 5
(SVT-like)



Condition 11
(SVT-like)

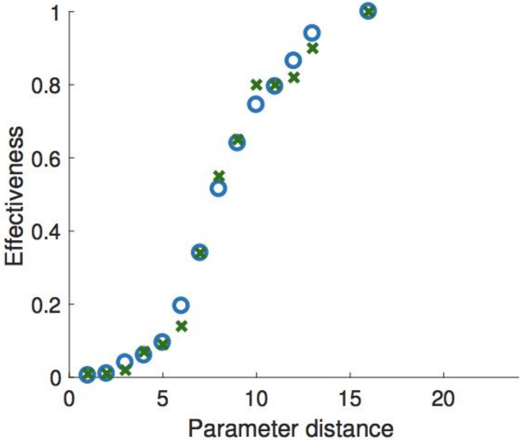
○ Training signals

✖ Validation signals

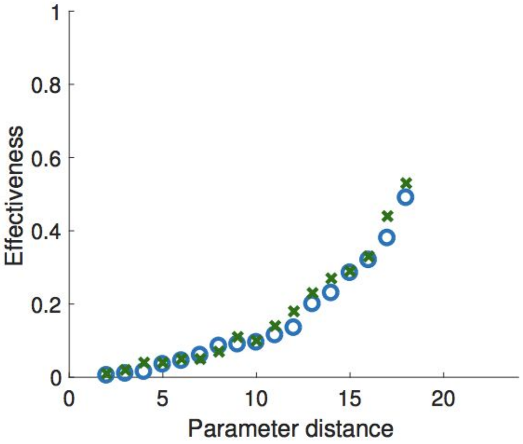
- Attacks on SVT-like conditions are not all equally effective
- Because, under normal HR, VT and VF must be reprogrammed to very low values to classify it as fast HR
- Common attack strategy: keep VF/VT thresholds and duration to a minimum

Evaluation, condition-agnostic attacks

- Two groups of signals obtained by merging VT-like and SVT-like EGMs
  - Useful when the attacker has little knowledge of the victim
  - 200 EGMs for training, 100 EGMs for validation



VT-like conditions



SVT-like conditions

Conclusion

- Attacks on cardiac devices are a serious threat, see previous studies and device recalls by FDA
- We presented the first method to synthesize stealthy reprogramming attacks tailored to the victim’s conditions
- Employs synthetic EGMs and automated reasoning (OMT) to find malicious parameters with optimal effectiveness-stealthiness trade-offs
- Well generalizes to unseen data (mimicking unknown victim EGM)
- **Future work:** other ICD models, real patient EGMs, closed-loop interaction, spoofing attacks