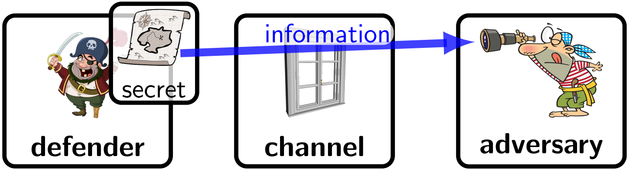
Adversary Gain vs. Defender Loss in Quantifying Information Flow

Piotr (Peter) Mardziel,[†] Mário S. Alvim,[+] Michael Hicks,[†]

[†]University of Maryland, College Park,
[+]Universidade Federal de Minas Gerais

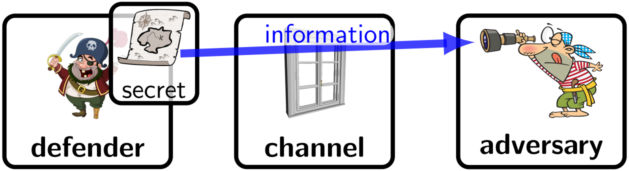
Quantified Information Flow [QIF]

- Secrets leak to bad guys.
- **Quantify leakage of the secret.**



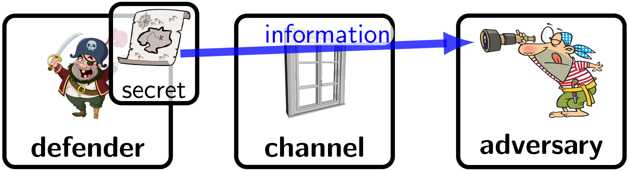
Why Quantified Information Flow?

- Evaluate risks.
- Evaluate relative merits of protection mechanisms.
- Design incentives to keep adversaries from participating.



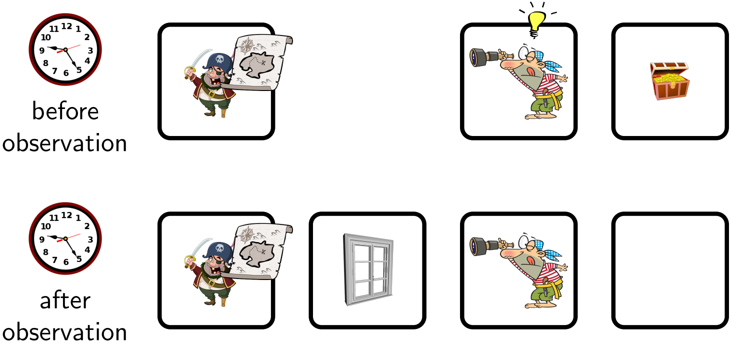
Examples

- Password authentication
- Location-based services
- Address space randomization



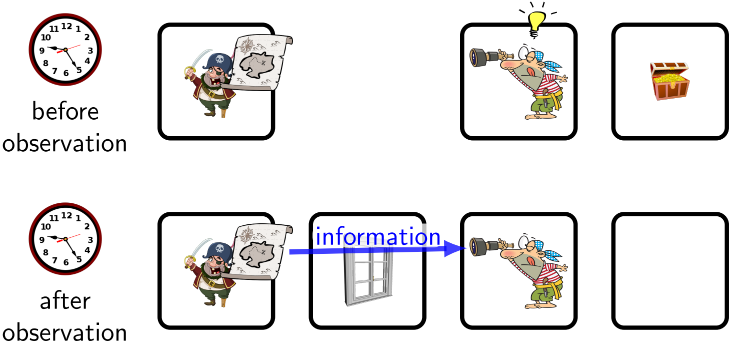
Quantified Information Flow: The Approach

- **Flow** $\stackrel{\text{def}}{=}$ increase in adversary's expected success
  - Model channel.
  - Model adversary behavior, exploitation.
  - Quantify expected success of optimal adversary.



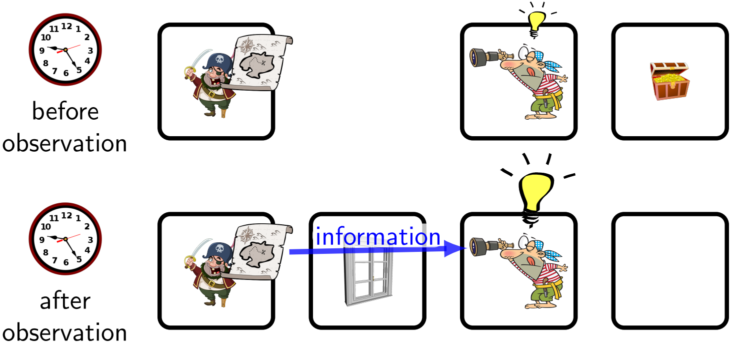
Quantified Information Flow: The Approach

- **Flow** $\stackrel{\text{def}}{=}$ increase in adversary's expected success
  - Model channel.
  - Model adversary behavior, exploitation.
  - Quantify expected success of optimal adversary.



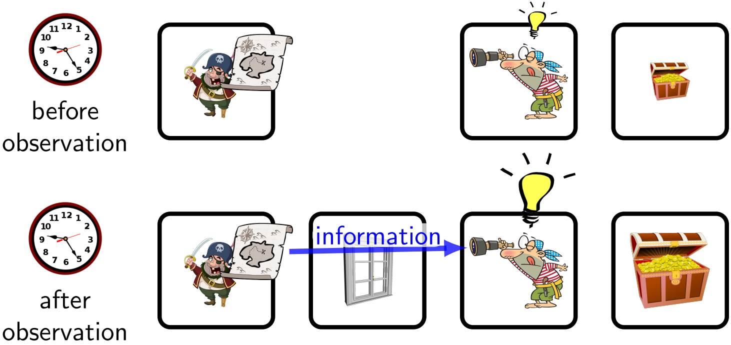
Quantified Information Flow: The Approach

- **Flow** $\stackrel{\text{def}}{=}$ increase in adversary's expected success
  - Model channel.
  - Model adversary behavior, exploitation.
  - Quantify expected success of optimal adversary.



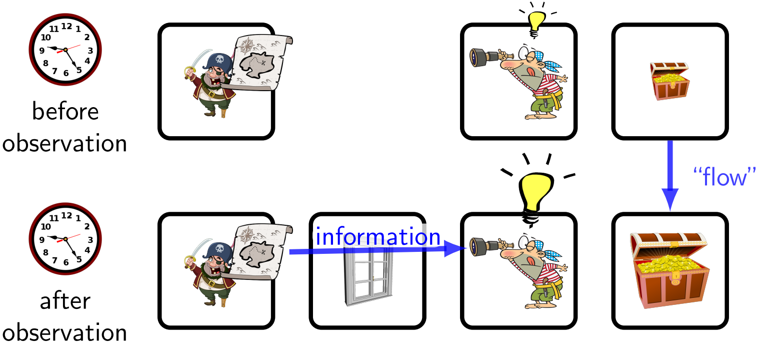
Quantified Information Flow: The Approach

- **Flow** $\stackrel{\text{def}}{=}$ increase in adversary's expected success
  - Model channel.
  - Model adversary behavior, exploitation.
  - Quantify expected success of optimal adversary.



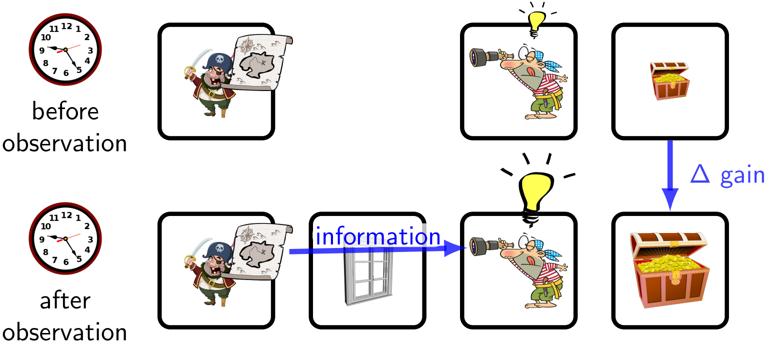
Quantified Information Flow: The Approach

- **Flow** $\stackrel{\text{def}}{=}$ increase in adversary's expected success
  - Model channel.
  - Model adversary behavior, exploitation.
  - Quantify expected success of optimal adversary.



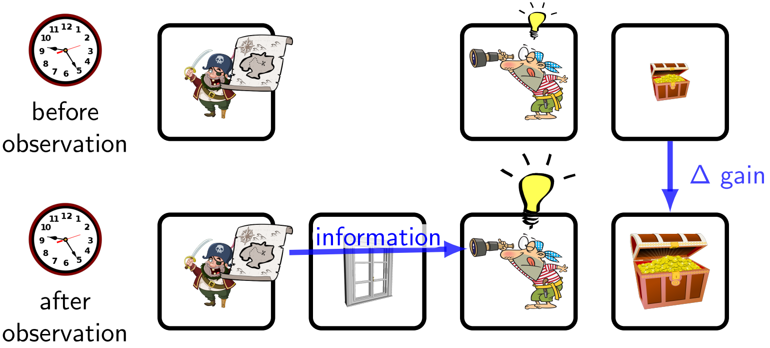
Quantified **Gain**: The Approach

- Δ adversary's **gain**
  - Model channel.
  - Model adversary behavior, exploitation.
  - Quantify (optimal adversary) **gain**.



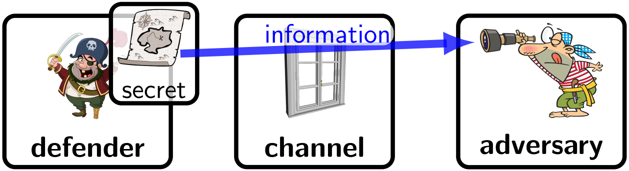
Quantified **Gain**: The Approach

- Δ adversary's **gain** $\stackrel{?}{=}$ Δ defender's **loss**
  - Model channel.
  - Model adversary behavior, exploitation.
  - Quantify (optimal adversary) **gain**.



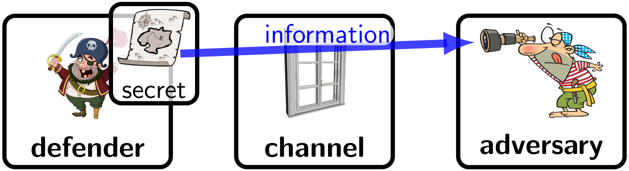
Examples

- Password authentication
- Location-based services
- Address space randomization



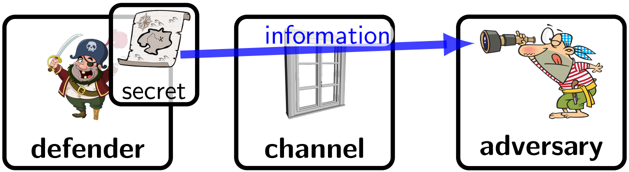
Examples

- Password authentication
  - Loss of bank contents $=$ gain of bank contents
  - Loss of private info $\lesssim$ gain (theft) of identity
- Location-based services
- Address space randomization



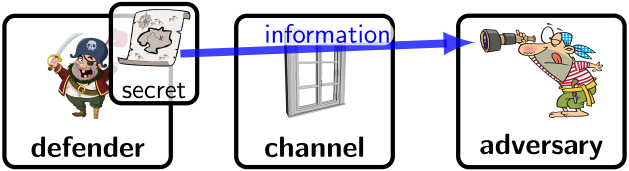
Examples

- Password authentication
  - Loss of bank contents $=$ gain of bank contents
  - Loss of private info $\lesseqgtr$ gain (theft) of identity
- Location-based services
  - Loss of privacy $\lesseqgtr$ Gain in targeted advertising
  - Loss of house contents $>$ Gain of stolen items
- Address space randomization



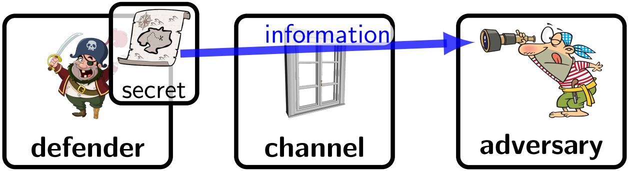
Examples

- Password authentication
  - Loss of bank contents $=$ gain of bank contents
  - Loss of private info $\lessgtr$ gain (theft) of identity
- Location-based services
  - Loss of privacy $\lessgtr$ Gain in targeted advertising
  - Loss of house contents $>$ Gain of stolen items
- Address space randomization
  - (Small) loss of service $<$ Gain of spam machine
  - Loss of critical service $>>$ Gain of spam machine



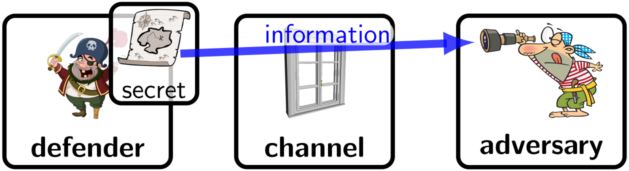
Examples

- Password authentication
  - Loss of bank contents $=$ gain of bank contents
  - Loss of private info $\lesseqgtr$ gain (theft) of identity
- Location-based services
  - Loss of privacy $\lesseqgtr$ Gain in targeted advertising
  - Loss of house contents $>$ Gain of stolen items
- Address space randomization
  - (Small) loss of service $<$ Gain of spam machine
  - Loss of critical service $>>$ Gain of spam machine
- ... depends



Examples

- Password authentication
  - Loss of bank contents $=$ gain of bank contents
  - Loss of private info $\lesssim$ gain (theft) of identity
- Location-based services
  - Loss of privacy $\lesssim$ Gain in targeted advertising
  - Loss of house contents $>$ Gain of stolen items
- Address space randomization
  - (Small) loss of service $<$ Gain of spam machine
  - Loss of critical service $>>$ Gain of spam machine
- ... depends
- **loss $\neq$ gain**



This work: Defender loss $\neq$ Adversary gain

$<$
$=$
$>$

- Defined model for information release with **distinct defender loss and adversary gain**.
- Both gain and loss are necessary to accurately quantify defender loss.
- Consequences about approximation:
  - Over-approximating adversary gain can be unsound
  - Over-approximating the channel (via partition refinement) can be unsound
- Worst-case (or best-case) metric to quantify the effect of catastrophic (or fortunate) defender behavior.

Outline

$<$

$=$

$>$

- Defined model for information release with **distinct defender loss and adversary gain**.
- Both gain and loss are necessary to accurately quantify defender loss.
- Consequences about approximation:
  - Over-approximating adversary gain can be unsound
  - Over-approximating the channel (via partition refinement) can be unsound
- Worst-case (or best-case) metric to quantify the effect of catastrophic (or fortunate) defender behavior.

Caveat

$<$

$=$

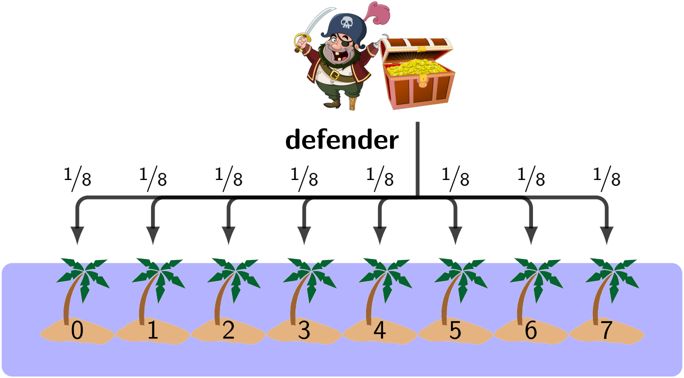
$>$

- Determining gain / loss is in the real world is hard.
- We assume the “instantaneous” gain and loss are given.
  - Gain and loss functions, *Secrets* $\times$ *Exploits* $\to \mathbb{R}$
- This work: analyze gain and loss dynamics as the adversary learns about the secret through some channel.

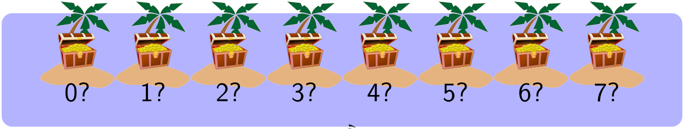
Example: Pirate Treasure

- Defender's reasoning about the adversary stealing his treasure and how to prevent it.
- Approximately password authentication.

Example: Pirate Treasure

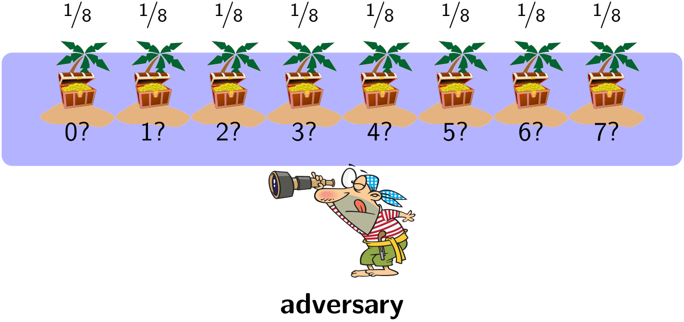


Secret Prior



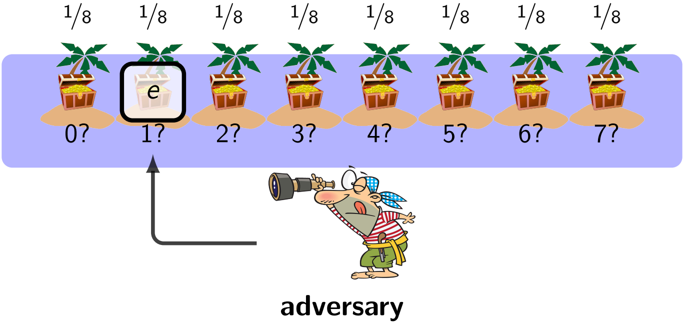
adversary

Secret Prior = Defender Belief



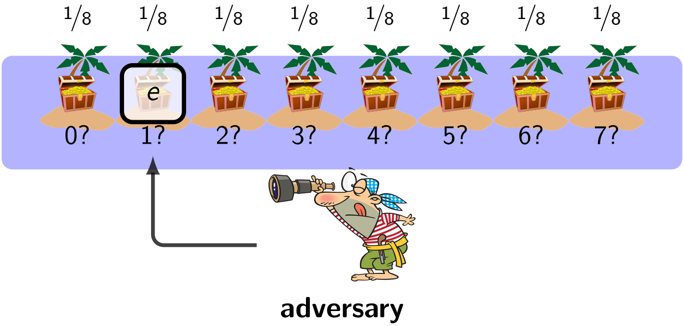
- Assume **adversary knows defender behavior**.

Exploitation



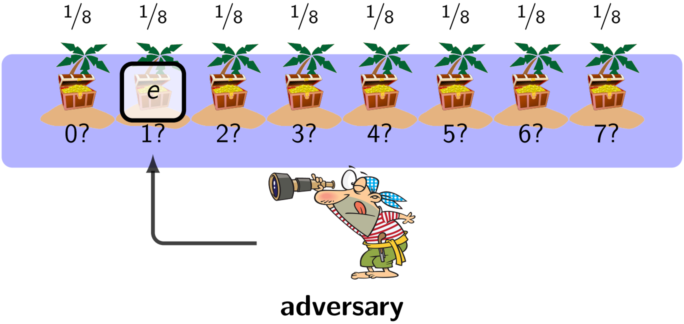
- Adversary “raids” an island e for the treasure. If $e = h$ he succeeds.

Exploitation



- Smith (FoSSaCS '09): (prior) **Vulnerability**: expected probability of optimal adversary with one guess being correct.

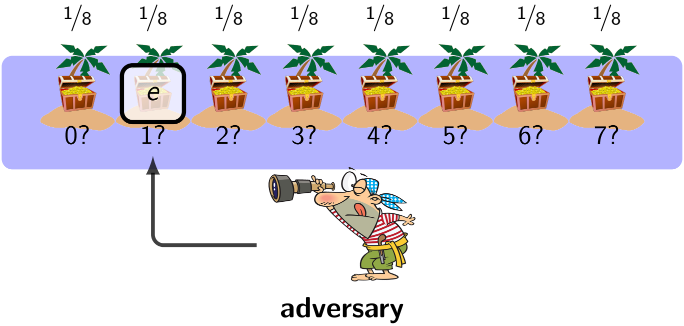
Exploitation: Measures of Success



Optimal adversary behavior:

- **Guessing Entropy**: Minimal number of guesses to find secret.
- Alvim et al. (CSF '12): g-**Vulnerability** Gain/payoff according to function $g(\mathrm{secret}, \mathrm{exploit})$.

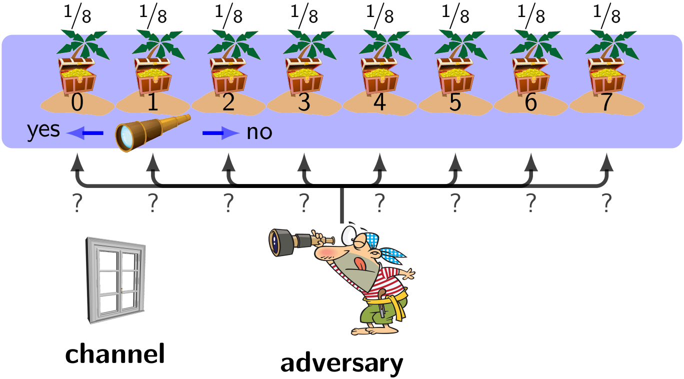
Exploitation: Vulnerability



- Connect probability of success to economic quantities.
- If the treasure is worth w doubloons, the expected gain to adversary and loss to the defender is $w \times \mathbb{V}$ doubloons. Here, $^{w}/_{8}$.
- Will stick with expected probability of success using the term "gain" in the remainder of this talk.

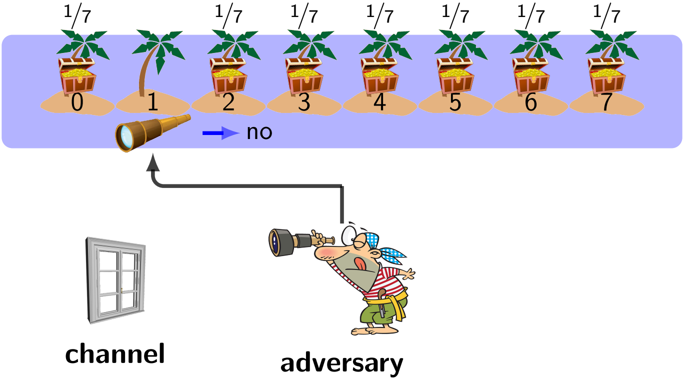
Observation

- Adversary can "stake out" an island to check whether the treasure is there.



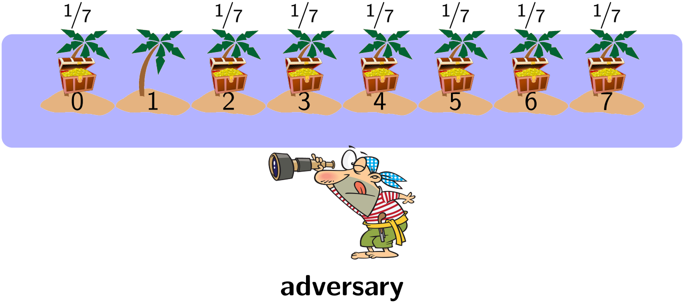
Observation

- Adversary can "stake out" an island to check whether the treasure is there.



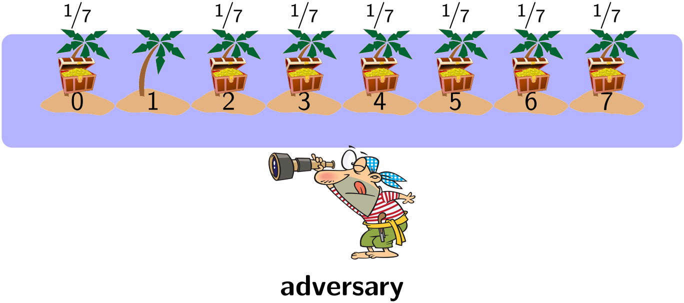
Increased knowledge

- Observation leads to increase in knowledge.
- Which leads to increased odds of exploitation.



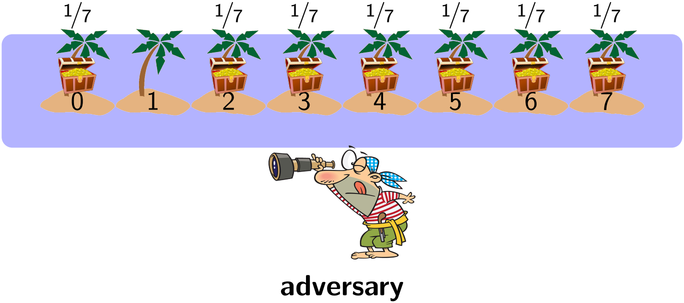
Increased knowledge ⇒ increased gain

- (posterior) **Gain**: expected probability of optimal adversary succeeding in one guess **given observation(s)**.
  - Optimal island to stake out.
  - Optimal island to raid.



Increased knowledge ⇒ increased gain

- (posterior) **Gain**: expected probability of optimal adversary succeeding in one guess **given observation(s)**.
  - Optimal island to stake out.
  - Optimal island to raid.
- Here: $\boxed{{}^1\!/_7}$.

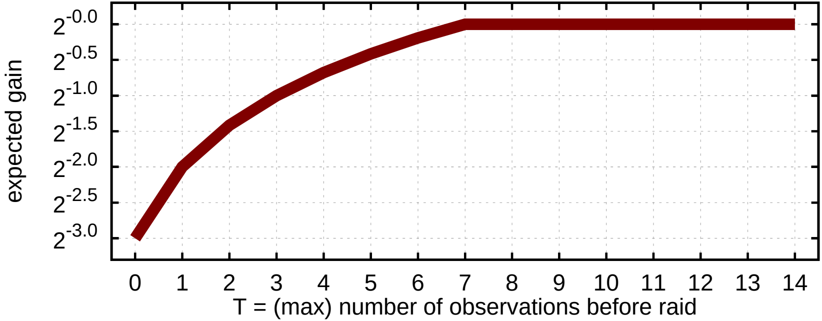


Observations over time

- Adversary continues observations (stake outs) but only has one exploitation chance (raid).
- How does their expected gain grow when they have more time to make observations?

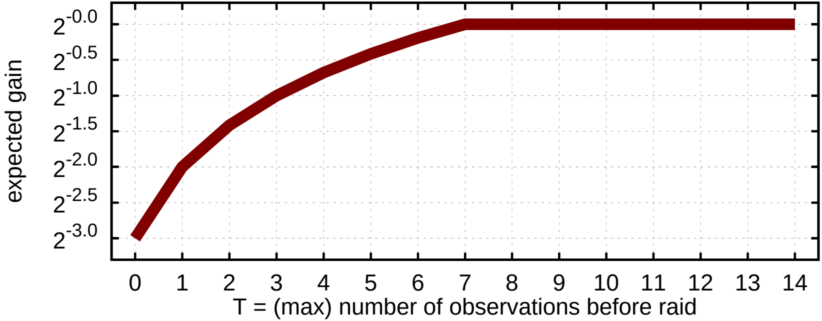
Observations over time

- Adversary continues observations (stake outs) but only has one exploitation chance (raid).
- How does their expected gain grow when they have more time to make observations?



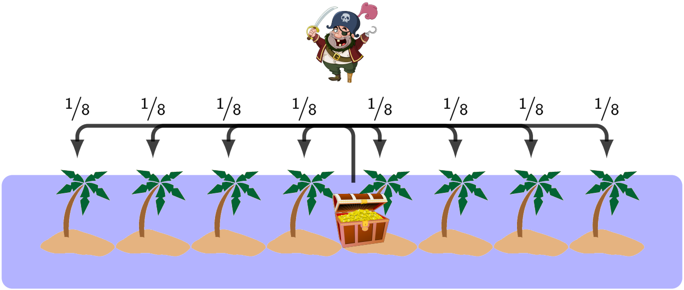
Observations over time

- Adversary continues observations (stake outs) but only has one exploitation chance (raid).
- How does their expected gain grow when they have more time to make observations?
- Eventually the treasure will be lost.



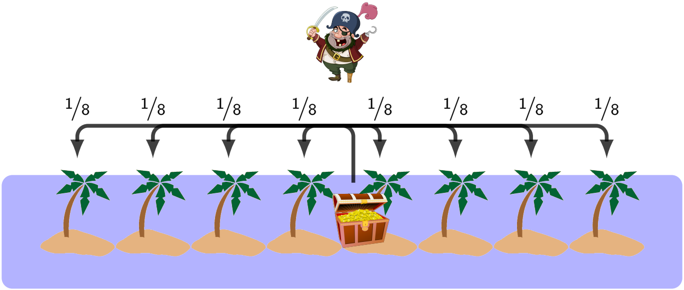
Moving the treasure

- Defender moves the treasure every once in a while.



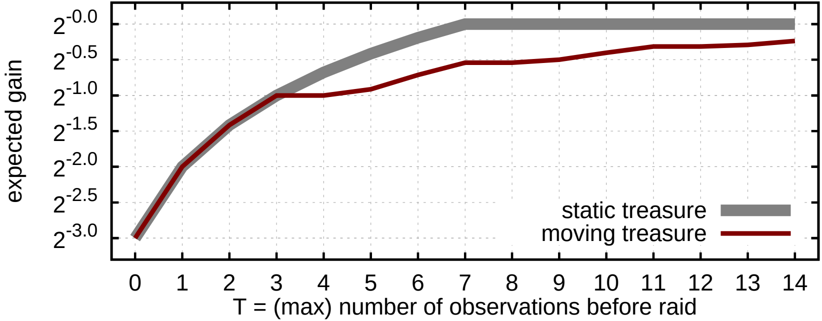
Moving the treasure

- Defender moves the treasure every once in a while.
- **Assume** adversary knows the process with which the defender does this.



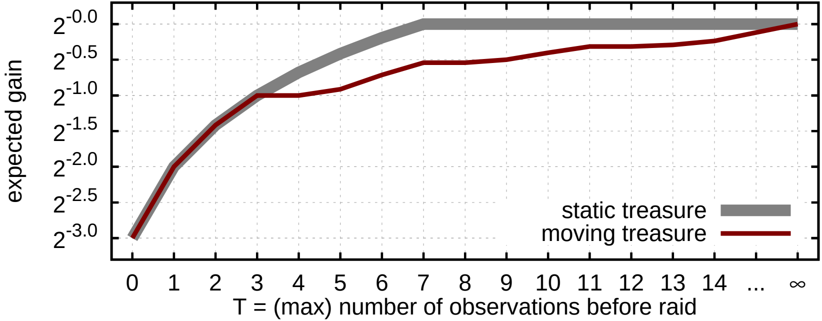
Gain with moving treasure

- Defender moves his treasure every 3 time steps.



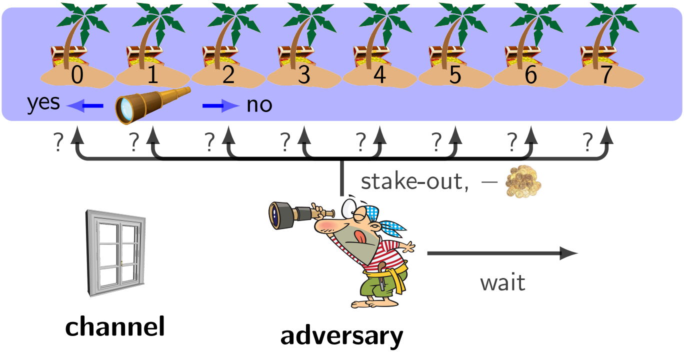
Gain with moving treasure

- Defender moves his treasure every 3 time steps.
- Eventually the treasure is lost.



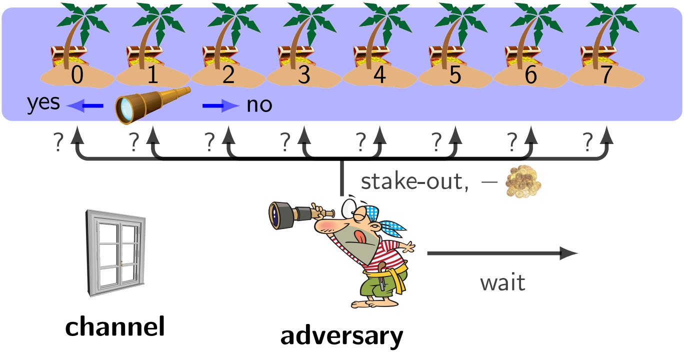
Costly Observation

- Defender makes it harder for the adversary to stake out for the treasure.
- It costs 0.10 [treasure] to stake out an island.



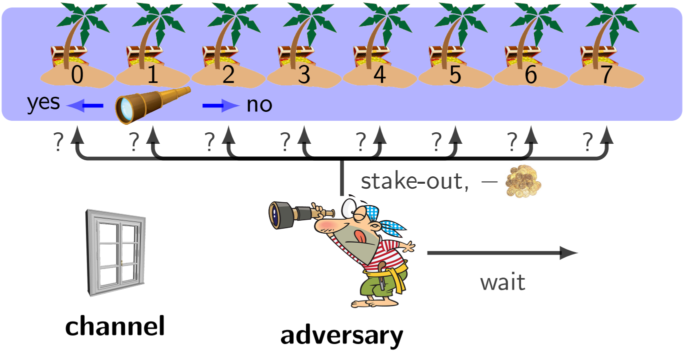
Costly Observation

- Defender makes it harder for the adversary to stake out for the treasure.
- It costs 0.10 [treasure] to stake out an island.
- Gain = (1.0 if treasure raided) - (0.1 * num. of observations)



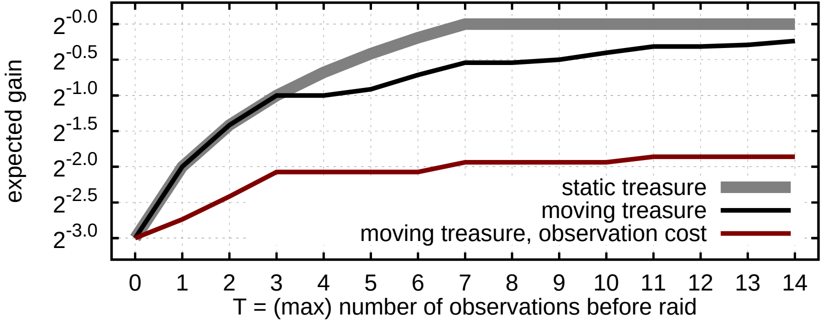
Costly Observation

- Defender makes it harder for the adversary to stake out for the treasure.
- It costs 0.10 [treasure] to stake out an island.
- Gain = (1.0 if treasure raided) - (0.1 * num. of observations)
  - **Gain can no longer be interpreted as chances of adversary successfully capturing the treasure.**



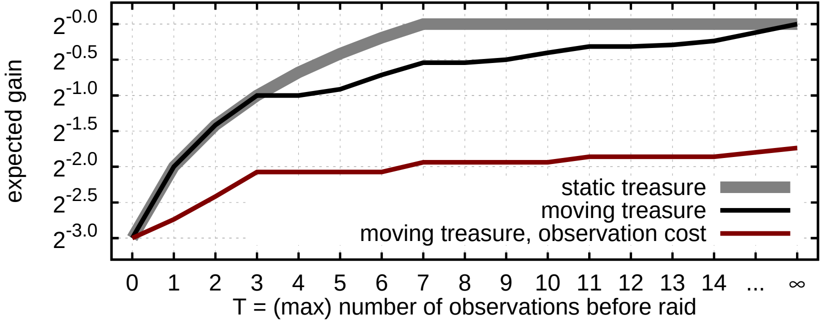
Gain with costly stake outs

- Defender still moves his treasure every 3 time steps.



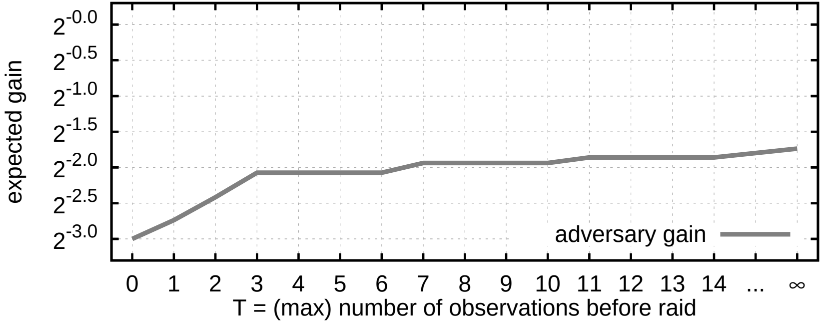
Gain with costly stake outs

- Defender still moves his treasure every 3 time steps.
- Adversary gain bounded even in the limit.



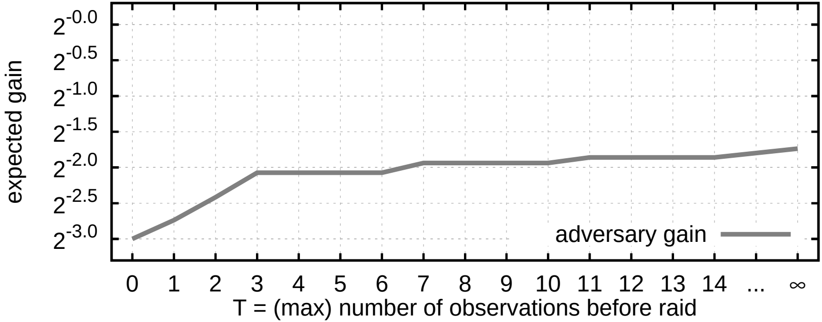
Gain with costly stake outs

- Defender wants $\leq 50\%$ chance of losing his treasure.
- Should he be satisfied with this result?



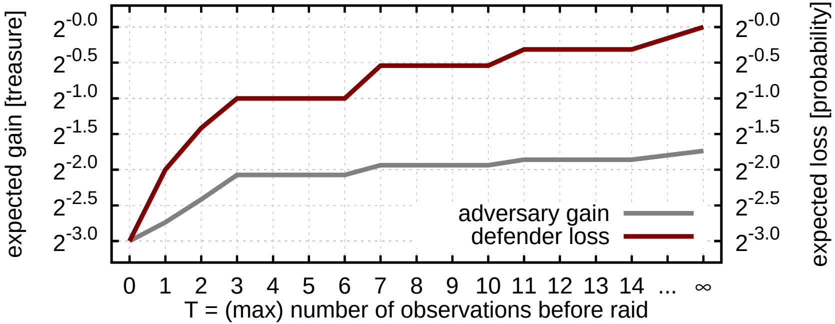
Gain with costly stake outs

- Defender wants $\leq$ 50% chance of losing his treasure.
- Should he be satisfied with this result?
- **Adversary gain does not measure defender loss.**



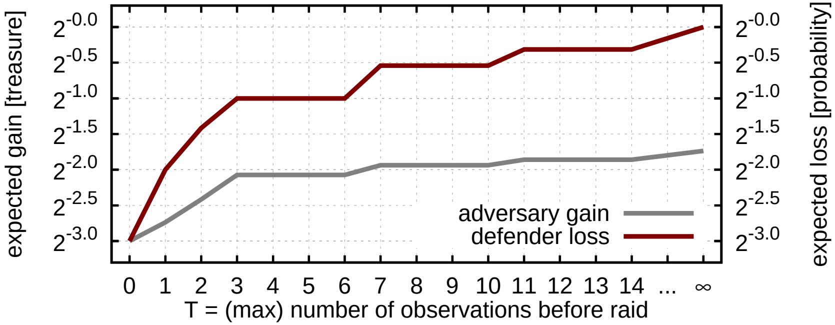
Gain with costly stake outs

- Defender wants $\leq 50\%$ chance of losing his treasure.
- Should he be satisfied with this result?
- **Adversary gain does not measure defender loss.**



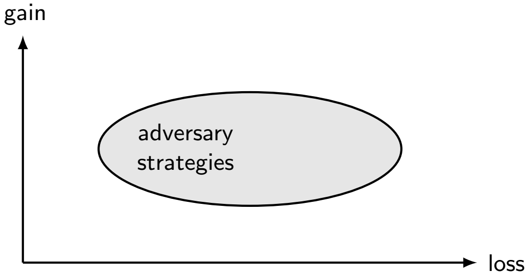
Gain with costly stake outs

- Defender wants $\leq 50\%$ chance of losing his treasure.
- Should he be satisfied with this result?
- **Adversary gain does not measure defender loss.**
- Optimal adversary will keep on staking out indefinitely.



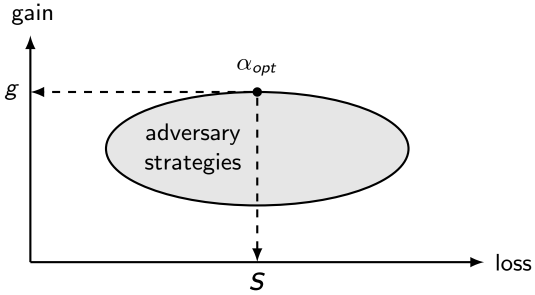
Dimensions of adversary strategy

- Each adversary strategy induces both a gain and a resulting defender loss.
  - Strategies: functions that determine adversary's action based on their past actions and observations.



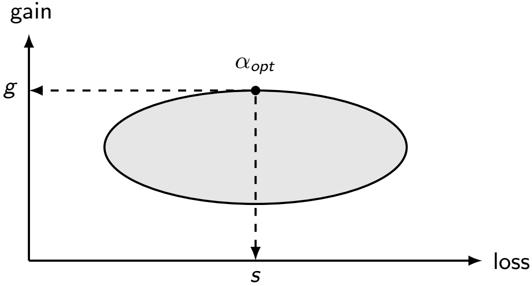
Dimensions of adversary strategy

- Each adversary strategy induces both a gain and a resulting defender loss.
  - Strategies: functions that determine adversary's action based on their past actions and observations.
- **Our metric: expected defender loss assuming adversary optimizes gain.**



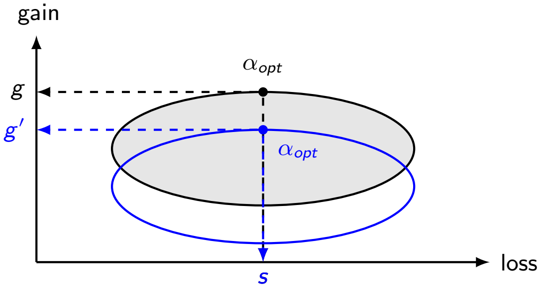
Gain + Loss

- **Cannot analyze a scenario in just one dimension.**



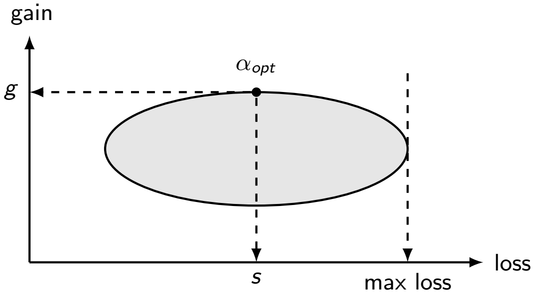
Gain + Loss

- **Cannot analyze a scenario in just one dimension.**
- Gain only: not what a defender is interested in.



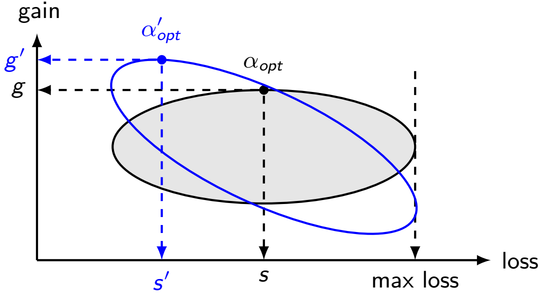
Gain + Loss

- **Cannot analyze a scenario in just one dimension.**
- Loss only: miss out on disincentives.



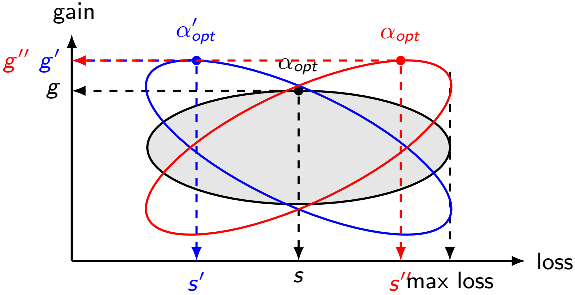
Gain + Loss

- **Cannot analyze a scenario in just one dimension.**
- Loss only: miss out on disincentives.
  - In example: stake out cost must be $\geq 1/7$ treasure units.

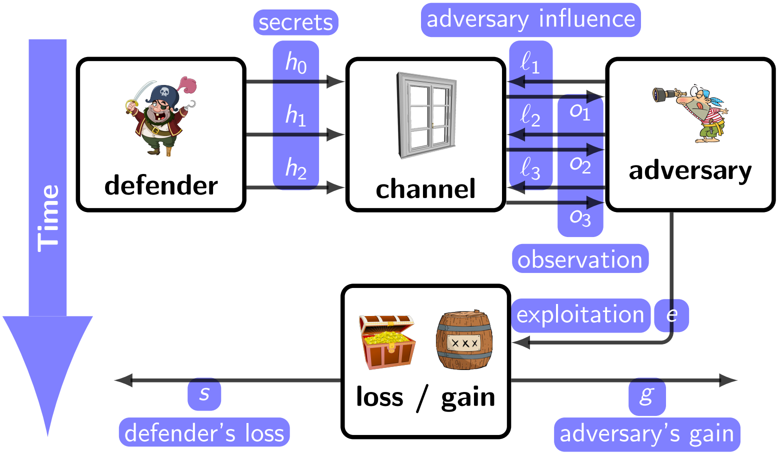


Gain + Loss

- **Cannot analyze a scenario in just one dimension.**
- Loss only: miss out on disincentives.
  - **Or miss bad incentives.**

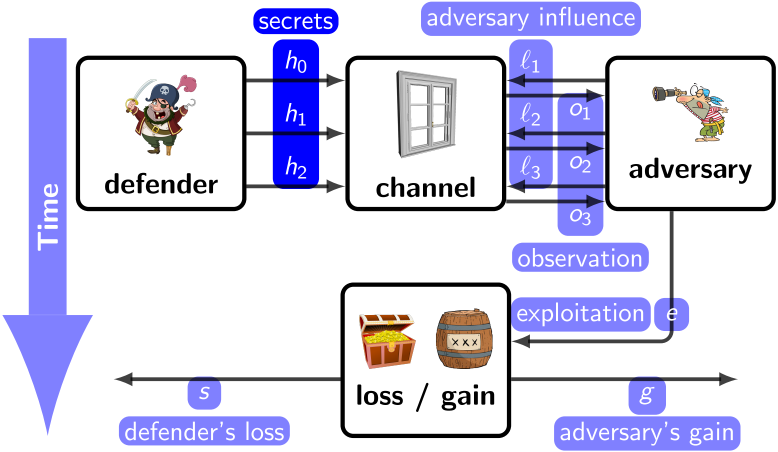


Model Overview



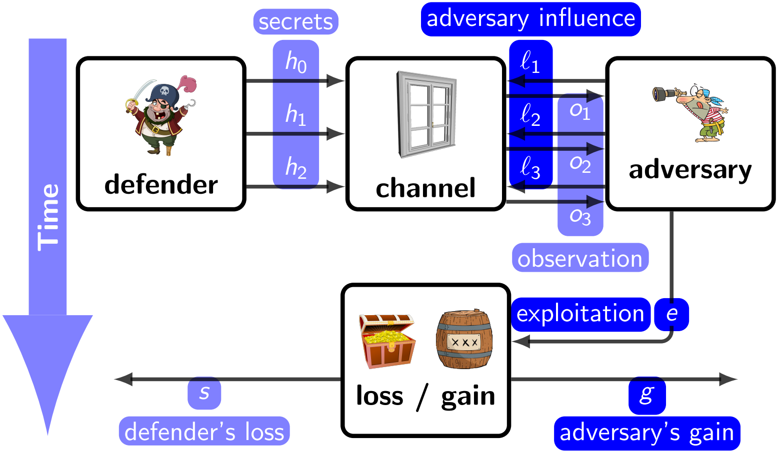
Model Overview

- Prior: initial secret distribution and distribution over (non-deterministic) functions describing secret evolution.



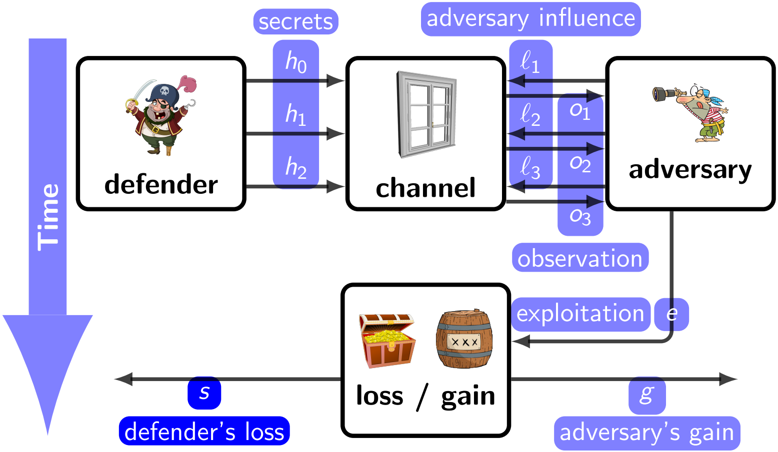
Model Overview

- Adversary optimization: optimize low inputs (channel influence) and exploitation for maximal gain.



Model Overview

- Measurement: measure the resulting defender loss.



Prototype, Prototyping Implementation

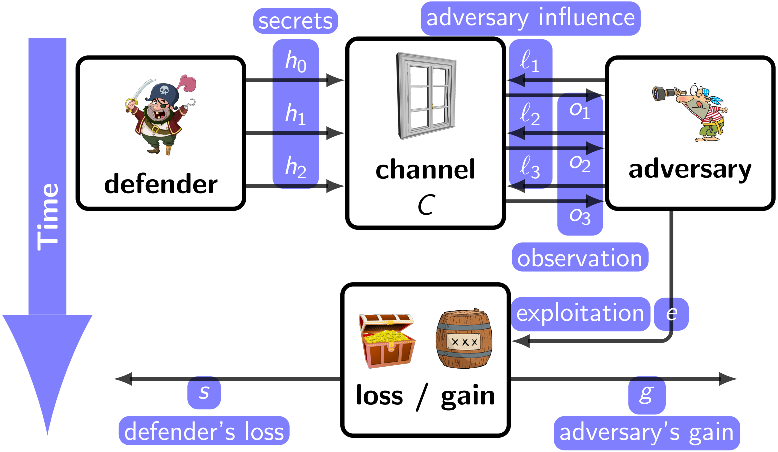
- Describe models as probabilistic programs in monadic-style OCaml.
- Optimize adversary behavior via backward induction.
- Compute the resulting defender loss.
- Analyze a series of scenarios (including this talk's examples)
- Freely available online.

Approximations

- Previously: both loss and gain are necessary.
- Previously: loss and gain functions might be hard to ascertain in the real world.
- Also: channel might be uncertain or too hard to analyze.
- Solution: over-approximations ?

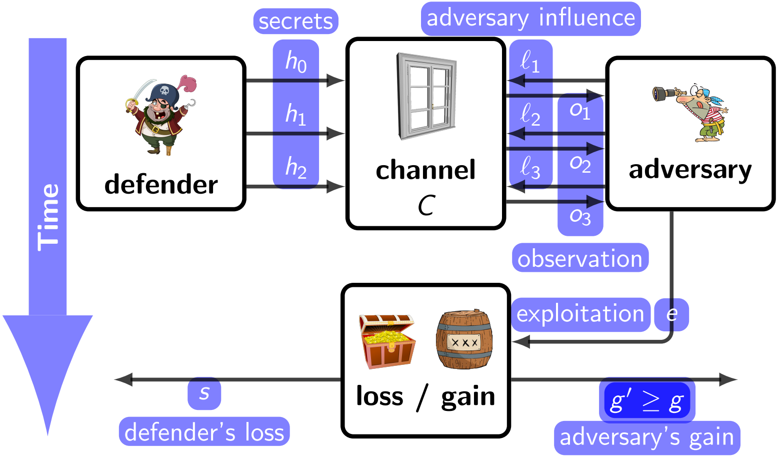
Soundness Approximations

- Does over-approximating gain or channel lead to over-approximation of loss?



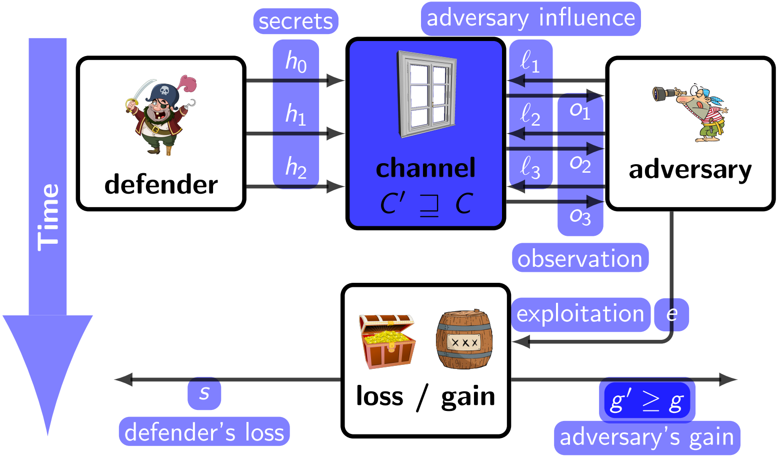
Soundness Approximations

- Does over-approximating gain or channel lead to over-approximation of loss?



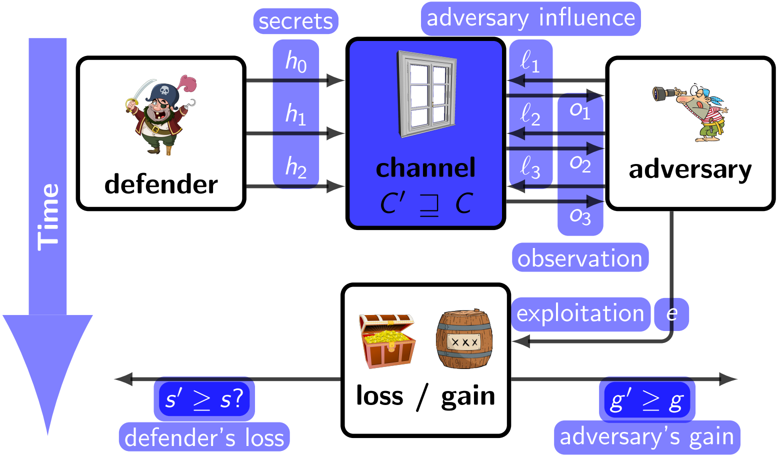
Soundness Approximations

- Does over-approximating gain or channel lead to over-approximation of loss?



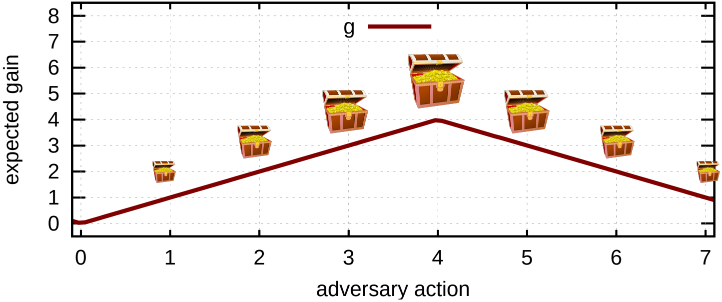
Soundness Approximations

- Does over-approximating gain or channel lead to over-approximation of loss?



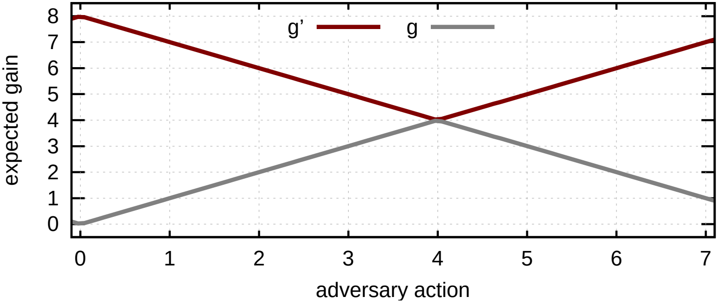
Approximating Gain

- Adversary gain (and defender loss): treasure is spread out around a central island; *secret* $= 4$ is shown below.



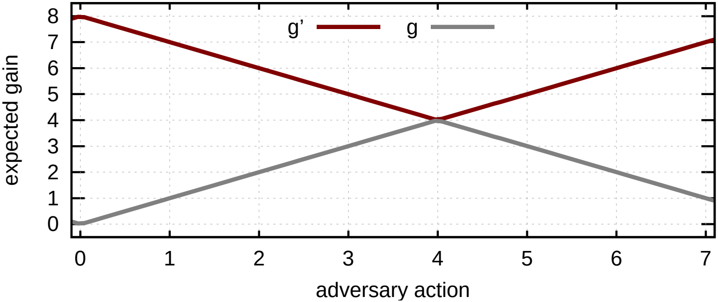
Approximating Gain

- Very bad over-approximation:
  $g'(secret, exploit) \geq g(secret, exploit)$



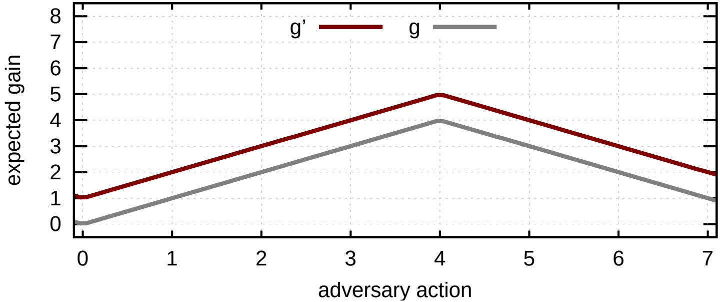
Approximating Gain

- Very bad over-approximation:
  $g'(secret, exploit) \geq g(secret, exploit)$
- **Not sound for loss.**



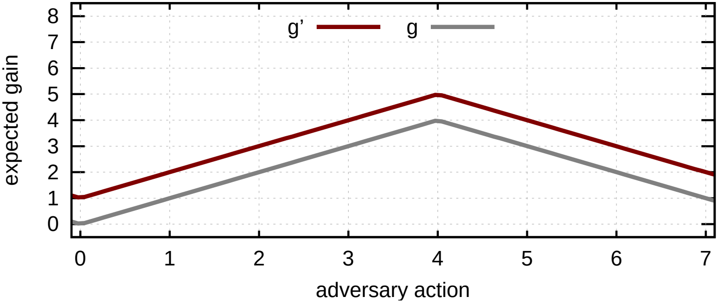
Approximating Gain

- Preference preserving approximation: $g(secret, exploit_1) \geq g(secret, exploit_2) \Leftrightarrow g'(secret, exploit_1) \geq g'(secret, exploit_2)$



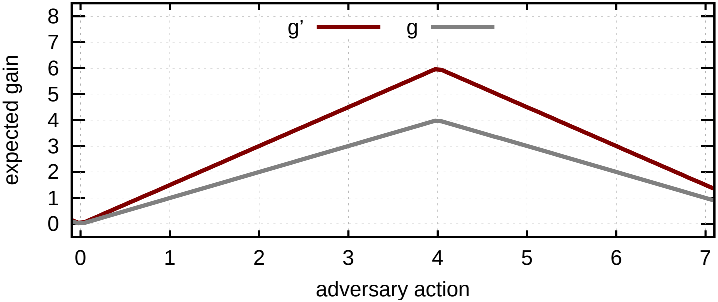
Approximating Gain

- Preference preserving approximation: $g(secret, exploit_1) \geq g(secret, exploit_2) \Leftrightarrow g'(secret, exploit_1) \geq g'(secret, exploit_2)$
- **Not sound for loss.**



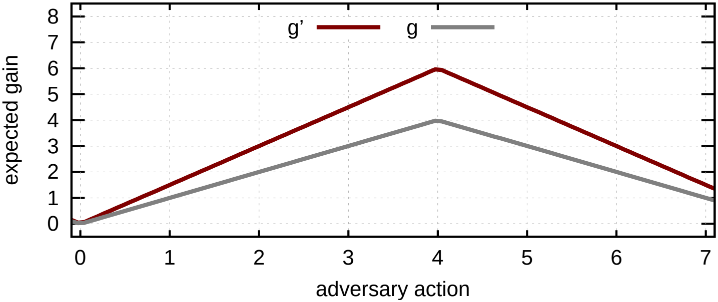
Approximating Gain

- Linear scaling: $g'(secret, exploit) = r * g(secret, exploit)$ with $r > 0$.



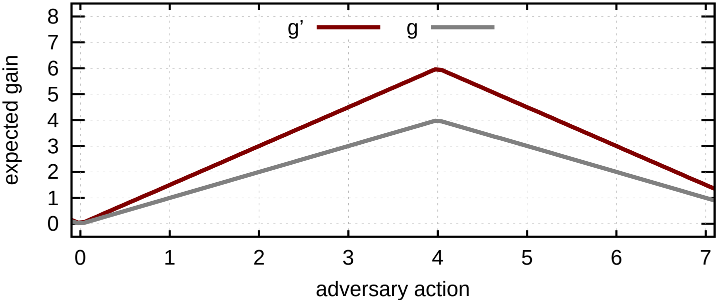
Approximating Gain

- Linear scaling: $g'(secret, exploit) = r * g(secret, exploit)$ with $r > 0$.
- Sound approximation, but (arguably) not useful.



Approximating Gain

- Linear scaling: $g'(secret, exploit) = r * g(secret, exploit)$ with $r > 0$.
- **(the only)** Sound approximation, but (arguably) not useful.



Approximating Channel

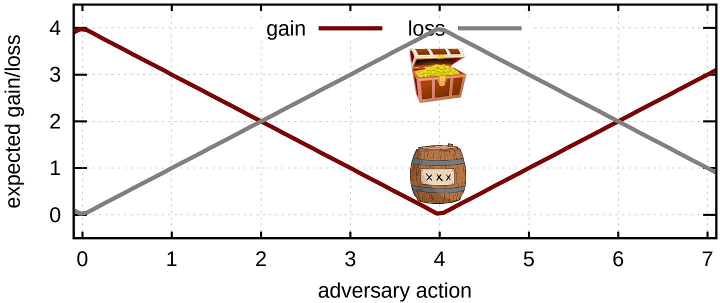
- Approximate channel C with $C' \sqsupseteq C$.
  - Example: C(secret) = "nothing" and C'(secret) = secret

Approximating Channel

- Approximate channel C with $C' \sqsupseteq C$.
  - Example: C(secret) = "nothing" and C'(secret) = secret
- **Not sound for loss.**

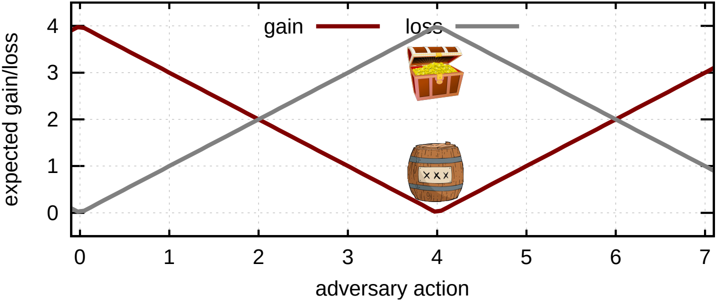
Approximating Channel

- Approximate channel C with $C' \sqsupseteq C$.
  - Example: C(secret) = "nothing" and C'(secret) = secret
- **Not sound for loss.**
- Example: Assume inverse relationship between gain and loss.



Approximating Channel

- Approximate channel C with $C' \sqsupseteq C$.
  - Example: C(secret) = "nothing" and C'(secret) = secret
- **Not sound for loss.**
- Example: Assume inverse relationship between gain and loss.
  - The more the adversary knows, the less loss is incurred by defender.



Soundness of Approximations

- No "useful" approximations of gain are sound for loss.
- Conjecture: no approximation of channel is sound for loss.

Conclusions

- Model for information flow distinct adversary gain and defender loss.

Conclusions

- Model for information flow distinct adversary gain and defender loss.
- Both gain and loss are necessary for accurate measurement of loss.

Conclusions

- Model for information flow distinct adversary gain and defender loss.
- Both gain and loss are necessary for accurate measurement of loss.
- Unsound consequences for loss when over-approximating gain or channel.

Conclusions




- Model for information flow distinct adversary gain and defender loss.
- Both gain and loss are necessary for accurate measurement of loss.
- Unsound consequences for loss when over-approximating gain or channel.
- Implementation and Experiments

Adversary Gain vs. Defender Loss in Quantified Information Flow

- Model for information flow distinct adversary gain and defender loss.
- Both gain and loss are necessary for accurate measurement of loss.
- Unsound consequences for loss when over-approximating gain or channel.
- Implementation and Experiments
- `http://ter.ps/fcs14`
  - This paper, Oakland'14 paper, TR, Implementation, Experiments