# Technical Notes

*TECHNICAL NOTES are short manuscripts describing new developments or important results of a preliminary nature. These Notes should not exceed 2500 words (where a figure or table counts as 200 words). Following informal review by the Editors, they may be published within a few months of the date of receipt. Style requirements are the same as for regular contributions (see inside back cover).*

# Monotone Optimal Threshold Feedback Policy for Sequential Weapon Target Assignment

Krishnamoorthy Kalyanam*
*Infoscitex Corporation, Dayton, Ohio 45431*
David Casbeer†
*U.S. Air Force Research Laboratory, Wright–Patterson Air Force Base, Ohio 45433*
and
Meir Pachter‡
*Air Force Institute of Technology, Wright–Patterson Air Force Base, Ohio 45433*

DOI: 10.2514/1.I010501

## I. Introduction

THE operational scenario is the following. A bomber with identical weapons travels along a designated route/path and sequentially encounters enemy (ground) targets. Should the bomber decide to engage a target, the target will be destroyed with a probability of $p < 1$. Upon successful elimination, the bomber receives a positive reward $r$ drawn from a fixed known distribution. We stipulate that, before engagement, the bomber observes the target and is made aware of the reward $r$. Furthermore, upon releasing a weapon, the bomber is alerted as to whether or not the deployed weapon was successful. In other words, we employ a shoot–look–shoot policy. If the target is destroyed, the bomber moves on to the next target. On the other hand, if the target was not destroyed, the bomber can either reengage the current target or move on to the next target in the sequence. The optimal closed-loop control policy that results in the maximal expected cumulative reward is obtained via stochastic dynamic programming. Not surprisingly, a weapon is dropped on a target if, and only if, the observed reward is no less than a stage- and state-dependent threshold value. We show that the threshold value, as a function, is monotonic decreasing in the number of weapons and monotonic nondecreasing in the number of targets left to be engaged.

## II. Weapon–Target Assignment Problem

It is clear that, if there is no feedback and if the reward values are known a priori, the problem collapses to a special case of Flood's static weapon–target assignment (WTA) problem (see [1]) and the optimal solution is obtained via the maximum marginal return algorithm (see [2]).

Moreover, if the weapons are not homogeneous and the kill probability varies with target type, the resulting static assignment problem is NP-complete (see [3]). Exact and heuristic algorithms to solve this version of the WTA problem were provided in [4–6]. An approximate algorithm for a dynamic WTA problem, wherein not all targets were known to the decision maker at the start of the mission, was provided in [7]. Decentralized cooperative control methods for a modified WTA problem, wherein weapons seek to achieve a prespecified probability of kill on each target, were proposed in [8]. An integrated problem of sensor management and WTA for missile defense was considered in [9]. For other prior work related to the dynamic WTA, we refer the reader to the survey paper in [10].

Our model embraces a shoot–look–shoot policy [11–15], in that homogeneous weapons are assigned one at a time with observations in between that assess the success or failure of the prior engagements. In a related work [16], Kalyanam et al. considered the scenario where the target rewards were deterministic and known before engagement. In sequential assignment problems of the kind considered herein, the decision rule usually takes the form of "attack the target if, and only if, its observed value $r$ is no less than a certain threshold $c$," where the optimal $c$ is to be determined. Moreover, intuition tells us that the optimal $c = c(t, k)$ should be monotonic increasing in $t$ and decreasing in $k$, where $t$ and $k$ are the number of remaining targets and weapons, respectively. If the probability of kill is $p = 1$ (i.e., a bomb dropped on a target alway s destroys it), there is no need for repeated engagement of the same target; and the resulting problem is similar to revenue management (RM), wherein the threshold monotonicity property is well known to hold (for details, see [17,18]). However, it is not obvious that the result holds for the case of $p < 1$, and we have established this result/generalization in this Note. We also point out that the case of $p < 1$ does not make sense in the context of RM because

Received 13 September 2016; revision received 22 December 2016; accepted for publication 24 December 2016; published online 15 February 2017.

*Research Scientist; krishnak@ucla.edu.

†Research Engineer, Autonomous Control Branch. Senior Member AIAA.

‡Professor, Electrical and Computer Engineering Department. Associate Fellow AIAA.

an offer once rejected therein cannot be revisited. However, it makes perfect sense in the case of a bomber/military scenario, where multiple weapons are frequently employed to destroy a target with observations made in between.

The model in [12] differs from ours, only in that the time between the appearance of targets therein is a random variable with an exponential distribution. We have simplified the setup, for the bomber scenario, where the time between visits to consecutive targets is fixed (flight time, perhaps) but otherwise irrelevant. This yields an elegant proof of the main monotonicity result, which is used in [12] without proof. It is also known that, if additional complicating factors such as a search cost for finding a target [13] or a scenario wherein weapons can be replenished [15] are considered, the threshold monotonicity property breaks down.

## III. Stochastic Dynamic Programming Model

We consider a dynamic variant of the WTA problem, wherein the targets are visited sequentially by the bomber. Furthermore, we also incorporate feedback, in that the bomber is informed about the success/failure of a weapon upon deployment. This allows for dynamic decision making, where a decision is made as to whether 1) an additional weapon is deployed on the current target if the previous engagement was unsuccessful or 2) the bomber moves on to the next target in the sequence.

Let $V(t,k|r)$ indicate the optimal cumulative reward (i.e., "payoff to go") that can be achieved when the bomber with $k > 0$ weapons arrives at the first of $t$ targets with an observed value $r$. Furthermore, let $W(t,k) = \boldsymbol{E}_x V(t,k|x)$ be the expected optimal cumulative reward from $t$ targets and $k$ weapons, before observation. It follows that $V(t,k|r)$ must satisfy the Bellman recursion:

$$V(t,k|r) = \max_{u=0,1}\{p(r + W(t-1,k-1)) + (1-p)V(t,k-1|r), W(t-1,k)\} \tag{1}$$

where the control action $u = 0, 1$ indicates whether the bomber should stay and deploy a weapon or simply move on to the next target. In Eq. (1), decision $u = 0$ results in the current target being destroyed with probability $p$, and $u = 1$ results in the bomber moving on to the next target in the sequence. If the current target is destroyed, the bomber receives an immediate reward of $r$. The corresponding optimal feedback policy $\mu(t,k|r)$ is therefore given by the maximizing control action in Eq. (1). If the bomber is at the last target and has $k > 0$ weapons at hand, the expected reward is given by the following:

$$\begin{aligned} V(1,k|r) &= r(1-q^k),\\ \Rightarrow W(1,k) &= \bar{r}(1-q^k), \qquad k > 0 \end{aligned} \tag{2}$$

where $q = 1 - p$. In other words, $q^k$ represents the probability that the last target is not destroyed by any of the $k$ weapons. So, $1 - q^k$ is the probability that it gets destroyed, thereby yielding the reward [Eq. (2)]. Here, $\bar{r}$ denotes the mean value of the reward distribution function. As mentioned earlier, the optimal policy has a special structure. Indeed, a weapon from an available inventory of $k$ weapons is dropped on a target of observed value $r$ if, and only if, $r$ exceeds a threshold or control limit $c(t,k)$. In the next section, we prove the main result that $c(t,k)$ is monotonic decreasing in $k$ and monotonic nondecreasing in $t$.

## IV. Monotone Threshold Policy

Let $\Delta_t(k) \coloneqq W(t,k+1) - W(t,k)$ indicate the expected marginal reward yielded by assigning an additional weapon over and above $k$ weapons to $t$ targets before observation.

*Proposition 1:* $\Delta_t(k)$ is a monotonic decreasing function of $k$.

We shall prove the preceding proposition later. Notice, however, that the marginal reward yielded by the last target in the sequence

$$\Delta_1(k) = W(1,k+1) - W(1,k) = p\bar{r}q^k \tag{3}$$

is clearly a decreasing function of $k$ given that $q < 1$.

Suppose Proposition 1 is true; i.e., $\Delta_{t-1}(k)$ is a monotonic decreasing function of $k$. Then, we can define $\kappa_t(r) = \min_{k=0,1,\ldots} k$ such that $pr \geq \Delta_{t-1}(k)$. Indeed, $\kappa_t(r)$ is the smallest nonnegative integer such that the immediate expected reward $pr$ is no smaller than the marginal expected reward $\Delta_{t-1}(k)$. With this definition, we show that a threshold policy is optimal.

*Lemma 1:* If $\Delta_{t-1}(k)$ is a monotonic decreasing function of $k$, the optimal policy is as follows:

$$\mu(t,k+1|r) = \begin{cases} 1, & k < \kappa_t(r),\\ 0, & \text{otherwise} \end{cases}$$

*Proof:* From the Bellman recursion [Eq. (1)], we have $V(t,k|r) \geq W(t-1,k)$. It follows that

$$p(r + W(t-1,k)) + qV(t,k|r) \geq pr + W(t-1,k) \geq W(t-1,k+1), \quad \forall\ k \geq \kappa_t(r) \tag{4}$$

where Eq. (4) follows from the definition of $\kappa_t(r)$. Recall the Bellman recursion [Eq. (1)]:

$$\begin{aligned} V(t,k+1|r) &= \max_{u=0,1}\{p(r + W(t-1,k)) + qV(t,k|r), W(t-1,k+1)\},\\ \Rightarrow V(t,k+1|r) &= p(r + W(t-1,k)) + qV(t,k|r), \quad \forall\ k \geq \kappa_t(r)\\ \Rightarrow \mu(t,k+1|r) &= 0, \qquad k \geq \kappa_t(r) \end{aligned} \tag{5}$$

We shall prove the second part of the result, i.e., $\mu(t,k+1) = 1, k < \kappa_t(r)$ by induction on $k$. Recall the definition of $\kappa_t(r)$, which gives us the following:

$$pr + W(t-1,k) < W(t-1,k+1), \quad \forall\ k < \kappa_t(r) \tag{6}$$

If $\kappa_t(r) = 0$, there is nothing left to prove. So, suppose $\kappa_t(r) > 0$. From the Bellman recursion [Eq. (1)], we have the following:

$$V(t,1|r) = \max_{u=0,1}\{pr, W(t-1,1)\} = W(t-1,1) \tag{7}$$

where Eq. (7) follows by applying Eq. (6) for the case of $k = 0$. So, we have $\mu(t,1|r) = 1$.

Suppose $V(t,h|r) = W(t-1,h)$ for some $h < \kappa_t(r)$. The Bellman recursion [Eq. (1)] yields the following:

$$\begin{aligned} V(t,h+1|r) &= \max_{u=0,1}\{p(r+W(t-1,h)) + qV(t,h|r), W(t-1,h+1)\},\\ &= \max_{u=0,1}\{pr + W(t-1,h), W(t-1,h+1)\},\\ &= W(t-1,h+1)\\ &\Rightarrow \mu(t,h+1|r) = 1 \end{aligned} \tag{8}$$

where Eq. (8) follows from applying Eq. (6) to the case of $k = h$. In summary, we have the following:

$$\mu(t,1|r) = 1 \quad \text{and} \quad \mu(t,h+1|r) = 1 \quad \text{if } \mu(t,h|r) = 1, \quad \text{for some } h < \kappa_t(r) \tag{9}$$

So, we conclude that

$$\mu(t,k+1|r) = 1 \quad \text{if } \mu(t,k+1|r) = 1, \quad \forall\ k < \kappa_t(r) \tag{10}$$

□

The preceding result tells us that one out of the current inventory of $(k+1)$ weapons is deployed on the current target of value $r$ if, and only if, the immediate expected reward $pr$ is no less than the marginal reward obtained by assigning an additional weapon over and above $k$ weapons to the $t-1$ remaining targets. Indeed, we have the following:

$$\mu(t,k+1|r) = \begin{cases} 1, & r < c(t,k)),\\ 0, & \text{otherwise} \end{cases}$$

The threshold value is given by

$$c(t,k) = \frac{1}{p}\Delta_{t-1}(k)$$

*Theorem 1:* $\Delta_t(k)$ is monotonic decreasing in $k$.

*Proof:* We prove the result by induction on the number of targets left $t$. From Eq. (3), we know that $\Delta_1(k)$ is monotonic decreasing in $k$. Let us suppose that $\Delta_{t-1}(k)$ is a decreasing function of $k$. Combining Eqs. (5) and (8), we can write the following:

$$V(t,k+1|r) = \begin{cases} W(t-1,k+1), & k < \kappa_t(r)\\ pr + W(t-1,k), & k = \kappa_t(r)\\ p(r+W(t-1,k)) + qV(t,k|r), & k > \kappa_t(r) \end{cases} \tag{11}$$

Let $\Gamma_t(k|r) = V(t,k+1|r) - V(t,k|r)$. So, we have the following:

$$\Gamma_t(k|r) = \begin{cases} \Delta_{t-1}(k), & k < \kappa_t(r),\\ pr, & k = \kappa_t(r) \end{cases} \tag{12}$$

For $\ell = k - \kappa_t(r) > 0$, we have by repeated application of Eq. (11):

$$\begin{aligned} V(t,k+1|r) &= pr\sum_{i=0}^{\ell} q^i + q^\ell W(t-1,\kappa_t(r)) + p\sum_{i=0}^{\ell-1} q^i W(t-1,k-i),\\ \Rightarrow \Gamma_t(k|r) &= p\sum_{i=0}^{\ell-1} q^i \Delta_{t-1}(k-i-1) + pq^\ell r \end{aligned} \tag{13}$$

We proceed to show that $\Gamma_t(k|r)$ as prescribed by Eqs. (12) and (13) is a decreasing function of $k$. By our induction argument, $\Gamma_t(k|r)$ decreases as $k$ goes from zero to $\kappa_t(r) - 1$. From the definition of $\kappa_t(r)$, we have $pr < \Delta_{t-1}(\kappa_t(r) - 1)$.

For any $\ell = k - \kappa_t(r) \geq 0$, using Eq. (13), we can write the following:

$$\Gamma_t(k+1|r) - \Gamma_t(k|r) = pq^\ell(\Delta_{t-1}(\kappa_t(r)) - pr) + p\sum_{i=0}^{\ell-1} q^i(\Delta_{t-1}(k-i) - \Delta_{t-1}(k-i-1)) < 0 \tag{14}$$

because $\Delta_{t-1}(k-i) < \Delta_{t-1}(k-i-1)$ per the induction argument, and $\Delta_{t-1}(\kappa_t(r)) \leq pr$ as per the definition of $\kappa_t(r)$. Hence, $\Gamma_t(k|r)$ is a strictly decreasing function of $k$. So, the expected marginal reward given by $\Delta_t(k) = \boldsymbol{E}_x\Gamma_t(k|x)$ is also a decreasing function of $k$. □

*Theorem 2:* $\Delta_t(k)$ is monotonic nondecreasing in the number of remaining targets $t$.

*Proof:* We shall show that, for a given $k$, $\Delta_t(k)$ is monotonic nondecreasing in $t$. For $k < \kappa_t(r)$, we have from Eq. (12)

$$\Gamma_t(k|r) = \Delta_{t-1}(k) \tag{15}$$

and, for $k = \kappa_t(r)$,

$$\Gamma_t(k|r) = pr \geq \Delta_{t-1}(k) \tag{16}$$

as per the definition of $\kappa_t(r)$. For $\ell = k - \kappa_t(r) > 0$, we have from Eq. (13)

$$\begin{aligned}\Gamma_t(k|r) &= p\sum_{i=0}^{\ell-1} q^i \Delta_{t-1}(k-i-1) + pq^\ell r \\ &> \Delta_{t-1}(k)p\sum_{i=0}^{\ell-1} q^i + pq^\ell r \end{aligned} \tag{17}$$

$$\begin{aligned} &= \Delta_{t-1}(k) + (pr - \Delta_{t-1}(k))q^\ell \\ &\Rightarrow \Gamma_t(k|r) > \Delta_{t-1}(k), \qquad k > \kappa_t(r) \end{aligned} \tag{18}$$

where Eq. (17) follows from the monotonicity result (Theorem 1) and Eq. (18) follows from the definition of $\kappa_t(r)$. It follows from Eqs. (15), (16), and (18) that $\Delta_t(k) = \boldsymbol{E}_x\Gamma_t(k|x) \geq \Delta_{t-1}(k)$. □

*Corollary 1:* Because

$$c(t,k) = \frac{1}{p}\Delta_{t-1}(k)$$

it is monotone decreasing in $k$ and monotone nondecreasing in $t$.

## V. Conclusions

A dynamic variant of the weapon–target assignment problem is considered, wherein targets are sequentially visited by a bomber equipped with homogeneous weapons with a probability of kill $p$. Feedback is available, and so the bomber is promptly informed about the failure or success of a deployed weapon. Stochastic dynamic programming yields the optimal policy, which specifies that a weapon is dropped if, and only if, the observed target value exceeds a threshold. The intuitive result that the threshold is monotonic decreasing in the number of weapons remaining and monotonic nondecreasing in the number of targets left to be engaged for the case of $p < 1$ is proven.

## References

[1] Manne, A. S., "A Target-Assignment Problem," *Operations Research*, Vol. 6, No. 3, 1958, pp. 346–351. doi:10.1287/opre.6.3.346

[2] denBroeder, G. G., Ellison, R. E., and Emerling, L., "On Optimum Target Assignments," *Operations Research*, Vol. 7, No. 3, 1959, pp. 322–326. doi:10.1287/opre.7.3.322

[3] Lloyd, S. P., and Witsenhausen, H. S., "Weapons Allocation Is NP-Complete," *Proceedings of the 1986 Summer Conference on Simulation*, Soc. for Computer Simulation, San Diego, CA, July 1986.

[4] Ahuja, R. K., Kumar, A., Jha, K. C., and Orlin, J. B., "Exact and Heuristic Algorithms for the Weapon-Target Assignment Problem," *Operations Research*, Vol. 55, No. 6, Nov.–Dec. 2007, pp. 1136–1146. doi:10.1287/opre.1070.0440

[5] Madni, A. M., and Andrecut, M., "Efficient Heuristic Approaches to the Weapon–Target Assignment Problem," *Journal of Aerospace Computing, Information, and Communication*, Vol. 6, No. 6, June 2009, pp. 405–414. doi:10.2514/1.34254

[6] Mekawey, H. I., EL-Wahab, M. S. A., and Hashem, M., "Novel Goal-Based Weapon Target Assignment Doctrine," *Journal of Aerospace Computing, Information, and Communication*, Vol. 6, No. 1, Jan. 2009, pp. 2–29. doi:10.2514/1.34374

[7] Murphey, R. A., "Approximation and Complexity in Numerical Optimization: Continuous and Discrete Problems," *Nonconvex Optimization and Its Applications*, An Approximate Algorithm for a Weapon Target Assignment Stochastic Program, Vol. 42, Springer, Boston, MA, 2000, pp. 406–421. doi:10.1007/978-1-4757-3145-3_24

[8] Volle, K., Rogers, J., and Brink, K., "Decentralized Cooperative Control Methods for the Modified Weapon–Target Assignment Problem," *Journal of Guidance, Control, and Dynamics*, Vol. 39, No. 9, Sept. 2016, pp. 1934–1948. doi:10.2514/1.G001752

[9] Ezra, K. L., DeLaurentis, D. A., Mockus, L., and Pekny, J. F., "Developing Mathematical Formulations for the Integrated Problem of Sensors, Weapons, and Targets," *Journal of Aerospace Information Systems*, Vol. 13, No. 5, May 2016, pp. 175–190. doi:10.2514/1.I010372

[10] Huaiping, C., Jingxu, L., Yingwu, C., and Hao, W., "Survey of the Research on Dynamic Weapon-Target Assignment Problem," *Journal of Systems Engineering and Electronics*, Vol. 17, No. 3, Sept. 2006, pp. 559–565. doi:10.1016/S1004-4132(06)60097-2

[11] Mastran, D. V., and Thomas, C. J., "Decision Rules for Attacking Targets of Oppurtunity," *Naval Research Logisitics*, Vol. 20, No. 4, Dec. 1973, pp. 661–672. doi:10.1002/nav.3800200407

[12] Kisi, T., "Suboptimal Decision Rule for Attacking Targets of Opportunity," *Naval Research Logisitics*, Vol. 23, No. 3, Sept. 1976, pp. 525–533. doi:10.1002/nav.3800230313

[13] Sato, M., "A Sequential Allocation Problem with Search Cost Where the Shoot-Look-Shoot Policy Is Employed," *Journal of the Operations Research Society of Japan*, Vol. 39, No. 3, Sept. 1996, pp. 435–454.

[14] Sato, M., “On Optimal Ammunition Usage When Hunting Fleeing Targets,” *Probability in the Engineering and Informational Sciences*, Vol. 11, No. 1, Jan. 1997, pp. 49–64.
doi:10.1017/S0269964800004678

[15] Sato, M., “A Stochastic Sequential Allocation Problem Where the Resources Can Be Replenished,” *Journal of the Operations Research Society of Japan*, Vol. 40, No. 2, June 1997, pp. 206–219.

[16] Kalyanam, K., Rathinam, S., Casbeer, D., and Pachter, M., “Optimal Threshold Policy for Sequential Weapon Target Assignment,” *20th IFAC Symposium on Automatic Control in Aerospace*, edited by de Lafontaine, J., IFAC-PapersOnLine, Vol. 49, Elsevier, Sherbrooke, QC, Canada, Aug. 2016, pp. 7–10.
doi:10.1016/j.ifacol.2016.09.002

[17] van Ryzin, G. J., and Talluri, K. T., *An Introduction to Revenue Management*, INFORMS TutORials in Operations Research, pp. 142–194, Chap. 6.
doi:10.1287/educ.1053.0019

[18] Aydin, S., Akçay, Y., and Karaesmen, F., “On the Structural Properties of a Discrete-Time Single Product Revenue Management Problem,” *Operations Research Letters*, Vol. 37, No. 4, July 2009, pp. 273–279.
doi:10.1016/j.orl.2009.03.001

M. J. Kochenderfer
*Associate Editor*