40

LA-UR- 01-1219

Title: **ANALYSIS OF DECISIONS IN MULTILATERAL ENGAGEMENTS**

Author(s): Gregory H. Canavan, P-DO

Submitted to: For discussions outside the Laboratory

Date: February 28, 2001

Los Alamos
NATIONAL LABORATORY

ANALYSIS OF DECISIONS IN MULTILATERAL ENGAGEMENTS

Gregory H. Canavan

Game theory is used to integrate strategic exchange models and costs and analyze engagements in which it is uncertain which side will strike first. That provides insight into strike incentives, preemption, and willingness to strike first in interaction and escalation. Strategic force reductions depend on uncertainty to reduce strike incentives. Trilateral interactions produce incentives to both strike and preempt that increase for smaller forces.

I. Introduction

An earlier paper derived a version of game theory applicable to the detection of strategic arms control compliance.[1] This note uses that framework to discuss situations in which it is uncertain which side will initiate conflict. In such situations, game theory provides insight into the reasons for strikes, preemption, and willingness to strike first. It illustrates the key role of uncertainty about which side might strike first, defines relevant, measurable decision variables that permit the quantitative assessment of strike incentives, and indicates how to reduce them. It shows that crises are generated internally by the attempts of independent players, none of which controls all decisions, to avoid the cost of being struck, which can force them to strike first. Uncertainty reduces these incentives by admixing large second strike costs, which offset the small first strike costs that produce the incentives. It is shown that military strikes can be treated by an extension of academic analyses. The resulting formalism is applicable to non-nuclear applications, so it can be used to analyze historical interactions to infer opponents' objectives.

Strategic force reductions introduce new elements; the main one is the dissimilar survivabilities of current strategic forces. The relative constancy of first strike costs combined with the sharp decrease of second strike costs at smaller forces make averaging less effective. That suggests that smaller, less vulnerable strategic forces are not necessarily more stable, although that result depends on the assumption that damage objectives do not change at low force levels.

The introduction of a third side does not strongly perturb the interaction between two strong ones; however, a strong side can generate a large first strike with a small fraction of its forces, while the weak one could only generate a small second strike with all of its forces. As a result, the strong side has incentive to strike the weak side when it can, and the weak side has an incentive to preempt the strong when it must. These dual incentives deter both the strong and weak sides from striking at intermediate force levels, but increase the strong side's incentive to strike when the weak side's forces are very small. At some level, these incentives could lead to strikes by the strong side, unless the strong side's damage incentives are diminished accordingly.

II. Exchange model, costs, and indices

Exchange models, strike optimization, conversion of damages into first and second strike costs, and conversion those costs into stability indices are derived in companion notes and summarized below.[2] The two sides in conflict are identified only as Unprime, U, and prime, P, in accord with the absence or presence of primes on the symbols for their forces, strikes, costs, and indices. Unprime has M vulnerable missiles with m warheads each and N survivable missiles with n warheads each for a total of W = mM + nN warheads. Prime has M' vulnerable missiles with m' = warheads and N' survivable missiles with n' warheads each for a total of W' = m'M' + n'N' warheads. If U strikes first, it delivers a fraction f of its W weapons on P's M' vulnerable missiles and the remaining fraction 1 – f in a first strike of magnitude

$$F = (1 - f)W, \qquad (1)$$

on military value targets, i.e, airfields, ports, and post-strike recovery assets. The allocation to missiles, f, the primary independent variable in determining U's optimal strike, is determined by nonlinear optimization. U's counterforce strike delivers an average of r = fW/M' warheads on each of P's vulnerable missiles, which gives them an average survival probability Q' = 1 – rp for r < 1, and $Q' \approx q^r$ for r > 1, where p = 1 - q is their single shot kill probability. Prime's second strike

$$S' = (Q'm'M' + n'N'). \qquad (2)$$

is delivered on value, as unused weapons are taken to have no residual value. The equations for P's first strike, F', and U's second strike, S, are obtained by conjugating (interchanging primed and unprimed symbols in) Eqns. (1) and (2).

First and second strikes are converted into costs through exponential approximations to the value of military value targets destroyed, assuming that each side has 1/k ~ 1/k' ~ 1,000 value targets. The cost of damage to self and incomplete damage to other are joined with weighting parameters L and L', which measure an attacker's relative preference for damaging the other and preventing damage to self. The cost of U striking first, C_1, thus has two terms

$$C_1 = C_{1s} + C_{1o} = (1 - e^{-kS'}) / (1 + L) + Le^{-k'F} / (1 + L) = (1 - e^{-kS'} + Le^{-k'F}) / (1 + L), \quad (3)$$

where C_{1s} is U's cost of damage to self and C_{1o} is the cost of U's incomplete realization of its damage objective with respect to its opponent P. C_{1s} results from P's second strike, so it involves the magnitude of P's strike S' and the size of U's target set, 1/k, and approaches zero for small retaliation S' << 1/k. C_{1o} is the cost to U of its incomplete first strike on P, which approaches zero for F >> 1/k' and L/(1 + L) for F << 1/k'. If neither side strikes, U does not suffer damage to self, but it does experience the cost x = L/(1 + L) of not realizing any part of its damage objective. The cost to U of striking second after P strikes first is

$$C_2 = C_{2s} + C_{2o} = (1 - e^{-kF'}) / (1 + L) + Le^{-k'S} / (1 + L) = (1 - e^{-kF'} + Le^{-k'S})/ (1 + L), \quad (4)$$

Where C_{2s} is the cost to U of P's first strike and C_{2o} is the cost of U's diminished second strike. C_{2s} is large for F' large and S small, typical conditions for second strikes. L measures U's preference for doing damaging to P relative to that for preventing damage to self; thus, it is a measure of U's aggressiveness. The damage to self and other are borne by different parties, so the principal justification for adding them is that the results of doing so are plausible and not overly sensitive to the value of L used. P's first and second strike costs are obtained by conjugating Eqns (3) and (4). P's cost of inaction is x' = L'/(1 + L'). The cost-ratio stability index for U is $I = C_1 / C_2$, and that for P is $I' = C_1' / C_2'$, so a consistent composite cost ratio index is their product $I \times I' = (C_1 / C_2)(C_1' / C_2')$.[3]

To illustrate the role of first strike costs in optimizing allocations and first strikes, let U and P have equal, modest forces (kW << 1), use Eq. (1) for F, the conjugate of Eq. (2) for S, substitute them into Eq. (3), expand exponentials, and differentiate with respect to f to find

$$f_{opt} = \ln(-L/m\ln q)(M/W\ln q), \quad (5)$$

which is the allocation that minimizes first strike costs. It depends directly on the fraction of vulnerable weapons, M/W, and logarithmically on the survival probability, q, damage preference, L, and weapons per vulnerable missile, m. The target sets k and k' cancel. For p = 0.8, L = 0.6, m = 2, and half vulnerable forces, $f_{opt} \approx 26\%$, so about a quarter of the attack is allocated to vulnerable missiles. For L = 0.3, f_{opt} increases to 0.37 as a less aggressive attacker attempts to damage limit more strongly. This analytic optimization can be extended to unequal and more complicated forces. It is reasonably accurate for kW ~ 1, but numerical solutions for f are used below.

III. Decision Analysis

The strategic game is defined by the graph in extended form shown in Table I, in which decisions are formulated in a forward manner and solved with a backward sweep.[4] Decision nodes are labeled from top to bottom and right to left, in accord with this solution process. At node 8 on the left, the two sides decide whether or not to engage. If they choose not to, the engagement terminates with expected costs to U and P of (x, x'). If they decide to engage, play moves to node 7, where—following a model due to Schelling[5]—Nature (N) decides whether U or P can strike first in an externally caused crisis. N chooses, with probability u, the upper branch, where U can strike first, and with probability 1 – u the lower branch where P can. The probability u plays a important role, which is discussed below. It could represent uncertainty about which side might first become aware of a crisis or which might be better prepared to take advantage of it. The two sides may know the distribution of u in advance, but will not know the result of N's random decision until the crisis. N choosing which side can strike first gives both players consistent information sets, so constructive solutions of the graph are Nash equilibria, i.e., each side's decisions at each node are the optimal responses to the other side's optimal decisions. Thus, it is not necessary to consider sub-optimal variants to have confidence that the results are robust.

Order of play. If N chooses U, the engagement moves to node 5 on the upper branch, where U can strike P. That advances play along the upper branch to node 1. There, P can strike back or not, which terminates that path. If U does not strike at node 5, the engagement advances to node 2, where P can strike (upper branch) or not (lower branch), after which U can retaliate or not. If N selects P at node 7, the engagement enters the lower branch and advances to node 6. There, P can strike or not. The former advances play to node 3, where U can restrike or not. The latter advances play to node 4, where U can strike or not. It is not always possible to solve Table I analytically, but it is always possible to solve it backward from the costs in the right column by allowing the appropriate decision maker to evaluate each node by minimizing its cost. At nodes 3, 4, and 5, U is the decision maker; at nodes 1, 2, and 6, P is. At node 7, Nature averages over the costs of nodes 5 and 6, to produce the expected costs at node 8 that U and P compare with the costs of inaction to determine whether to engage. If they decide to engage, their expected costs are determined at the outset and their optimal decisions are determined by retracing the steps of the backward solution.

Payoffs. The right column of Table I shows the costs of each path in terms of first and second strike costs defined above. The costs shown for node 1 are (C_1, C_2'), which are shown below to be the minimum of the costs of the two actions shown by the arrows to the right of the node. To reach node 1, U must have struck P first at node 5. If P follows the upper arrow and strikes back at node 1, one of the two engagements considered in conventional first strike analyses, which produces costs (C_1, C_2'). If P follows the lower arrow, i.e., does not strike back, its cost is

$$C_2' \approx (1 - e^{-k'F} + L'e^{-k0})/(1 + L') = C_{2s}' + L'/(1 + L') = C_{2s}' + x', \qquad (6)$$

which is the cost of damage to self, C_{2s}', plus P's unfulfilled damage objective, x'. As C_{2s}' does not vary with P's decision to strike back, the cost for not striking is greater than that for striking by $x'(1 - e^{-kS'})$. At node 1, P is the decision maker, so it chooses the minimum of C_2' from the upper arrow and C_{2s}' + x' from the lower. The former is smaller by $x'(1 - e^{-kS'})$, so P chooses the upper branch and retaliates. As P has no other use for its missiles, if it is struck, it strikes back to get some value for them. Thus, if U and P reach node 1, their expected costs are the ordered pair (C_1, C_2'). The same logic applies to node 3 with P and U interchanged, so its value is (C_2, C_1').

At node 2, P is the decision maker, and U did not struck at node 5. If P decides to strike (upper branch), U follows logic like P's at node 1 and strikes back, producing costs (C_2, C_1'). If P does not strike, both sides will have declined the opportunity to strike, so the engagement terminates with costs (x, x'). Neither will have sustained damage from strikes, but both will have incurred costs equal to their damage objectives. P makes its decision at node 2 by selecting the minimum of C_1' and x'. If P is non-aggressive, x' is small, so the lower branch is chosen, and the value of node 2 is (x, x'). If P is aggressive, C_1' can fall below x', P strikes, and the value of node 2 is (C_2, C_1'). The transition is the force level at which the C_1' curve intersects x'. This logic also holds at node 4 with P and U interchanged, so its value is (x, x') for $C_1 > x$, and (C_1, C_2') for $C_1 < x$.

Limiting example. For $x \approx 0$ and $x' \approx 1$, i.e., a non-aggressive U facing a very aggressive P, the costs at nodes 1 through 4 reduce by inspection to (C_1, C_2'), (C_2, C_1'), (C_2, C_1'), and (x, x'). At node 5, U chooses the minimum of C_1 and C_2, which is C_1, so U strikes first, which gives node 5 the value (C_1, C_2'). At node 6, P chooses the minimum of C_1' and x', which is C_1', so it strikes first, which gives node 6 the value (C_2, C_1'). At node 7, nature weights the costs of nodes 5 and 6 by the probability u to produce the expected costs

$$(7U, 7P) = u(C_1, C_2') + (1-u)(C_2, C_1') = [uC_1 + (1-u)C_2, uC_2' + (1-u)C_1']. \quad (7)$$

Whether U decides to participate is determined by the minimum of $uC_1 + (1-u)C_2$ and x. For $x \approx 0$, the minimum is x, so U would not choose to engage for any costs or u. P's decision is determined by the minimum of $uC_2' + (1-u)C_1'$ and x', which for $x' \approx 1$ is $uC_2' + (1-u)C_1$. Thus, P wishes to engage for any value of the costs and u. Since P engages, U must too. For less extreme damage objectives, averaging over u can mitigate this behavior.

IV. Equal Offensive Forces

Equal offensive forces offensive forces provide a transparent example of the application of the game theoretic methodology.[6] The forces used are post-START III offensive forces of Table II consisting of $M = M' = 500$ vulnerable missiles with equal numbers of weapons ($m = m'$) and $N = N' = 250$ survivable missiles with $n = n' = 2$ weapons each. The two sides engage by increasing their weapons per vulnerable missile equally. Such forces could be produced by a reduction to the post-START III force levels followed by a decision by each to increase the number of weapons on each vulnerable missile. Damage preferences of $L = L' = 0.6$ used, which imply damage objectives $x = x' \approx 0.38$, corresponding to moderately aggressive contestants. All missiles are assumed to have a single shot kill probability against vulnerable missiles of $p = 0.8$.

Figure 1 shows the optimal allocation of weapons to vulnerable missiles, which by symmetry is $f = f' \approx 30\%$ at $m = m' = 1$. The allocation increases to $\approx 40\%$ by $m = 2$, and is roughly constant thereafter. The number of weapons allocated to each vulnerable missile increases from ≈ 0.6 at $m = 1$ to ≈ 1.7 by $m = 3$, where the survival probability of vulnerable missiles falls below 10%, vulnerable missiles represent a sink of weapons, and second strikes are largely composed of survivable forces. Figure 2 shows that the first strike, F, increases with the number of weapons, while the second strike, S', falls slightly due to damage limiting. Figure 3 shows that the first strike cost, C_1, falls and the second strike cost, C_2, increases monotonically with m. Figure 4 shows the cost ratio C_1/C_2 used as a stability index in conventional analyses. It falls strongly and monotonically for $m > 1$, reaching ≈ 0.6 (composite ratio of ≈ 0.4) by $m = 3$—due in about equal parts to the decrease of C_1 and increase of C_2.

Results. Figure 5 shows the costs of the upper branch nodes 1 and 2 as functions of m' (= m). Node 1 is a strike by U followed by a second strike by P, which have costs to U and P of (1U, 1P) = (C_1, C_2'). At node 2, C_1' > x' for m = m' < 3.5, so neither side strikes, and their costs are (x, x'). But by m ≈ 3.5, C_1' falls to x' ≈ 0.38, so P has an incentives to strike. However, on the upper branch, only U can strike first. Recognizing that P will strike first at node 2 if given the option, U strikes first at node 5. Figure 6 shows the resulting costs, which are (x, x') for m < 3.5 and (C_1, C_2') for m > 3.5. At the transition at m = 3.5, C_1 = x, so there is no discontinuity in U's cost, but P's cost jumps from C_1' = x' to C_2' >> x'. So on the upper branch, P faces large cost increases due to decisions U makes on the basis of apparently small incremental costs. However, if U did not strike at node 5, the option to strike would pass to P at node 2, where P would strike for m' > 3.5 to realize its minimum cost C_1'. Then, P's cost would be continuous at C_1' = x', but U's would jump from x to C_2 >> C_1. The difference between the cost C_2 for preempting and C_1 for inaction gives an alternative explanation of U's incentive to strike first at node 5.

In terms of the decisions on Table I, as long as C_1' > x', P chooses inaction at node 2, which produces costs (x, x'). For equal forces, C_1 = C_1', so C_1' > x' implies C_1 > x, which means U's choice at node 5 is also inaction, so their costs are (x, x'). When C_1' falls below x', P would prefer to strike at node 2, which would produce costs (C_2, C_1'). Anticipating that shift, U would strike first at node 5, producing costs (C_1, C_2'). Thus, the actual decision variable is not the small difference C_1' – x' that P sees at node 2, but the much larger difference C_2 - C_1 that U would experience if it did not act at node 5. The other criteria for U to act is C_2 > C_1, which is generally met. The magnitude of the external event that triggers the crisis only enters this decision logic if it significantly alters P's damage objective x'. The assumption of equal forces is not important to U's decision to strike at node 5. The decision is, however, sensitive to U's inference that C_1' is approaching x'. U cannot determine either C_1' or x' precisely, so there is some danger that U will misjudge one or the other and strike either prematurely or too late.

By the symmetry of forces and objectives assumed, the costs for nodes 3, 4, and 6 are the same as those at nodes 1, 2, and 5 with U and P interchanged. Figure 7 shows P's costs at nodes 5, 6, and 7. The top curve is P's cost at node 5, where U strikes first. That corresponds to u = 1 in Eq. (7), for which for m > 3.5 the expected costs reduce to

$$(7U, 7P) = 1(C_1, C_2') + 0(C_2, C_1') = (C_1, C_2'). \qquad (8)$$

If U was sure it could strike first, it could reduce its cost by escalating to m > 3.5 and striking, which would greatly increase P's cost. The bottom curve on Fig. 7 is P's cost at node 6, where P strikes first. That corresponds to u = 0, which for m > 3.5

$$(7U, 7P) = 0(C_1, C_2') + 1(C_2, C_1') = (C_2, C_1'). \qquad (9)$$

If P was sure it could strike first, it could reduce costs. The middle curve corresponds to u = 0.5, i.e., equally aggressive competitors, for which the expected costs are

$$(7U, 7P) = 0.5(C_1, C_2') + 0.5(C_2, C_1') = [(C_1 + C_2)/2, (C_2' + C_1')/2], \qquad (10)$$

the average of each side's first and second strike costs. Figure 3 shows that for these symmetric conditions, U and P's second strike costs increase faster than their first strike costs fall, so their average is greater than (x, x'). Neither side sees an incentive to strike at node 7, so neither sees an incentive to engage at node 8. The admixture of large second strike costs offsets the small first strike costs that are the source of the strike incentives, which removes any incentive to strike. Given the uncertainty as to which side could strike first, neither sees an incentive to escalate to m > 3.5. Instead, both see an incentive to avoid that region, where uncertainty about which could strike first leads to expected costs higher than those of inaction. Stability obtains as long as u is within certain limits, which can be bounded by inverting Eq. (7), whose U component is

$$u = (C_2 - 7U) / (C_2 - C_1), \qquad (11)$$

which is the probability u that produces a value 7U for given costs C_1 and C_2. Substituting x for 7U gives u_U, the maximum value of u for which 7U > x, which assures that U does not see any strike incentives. Figure 3 shows that at m = 5, $C_1 \approx 0.33$ and $C_2 \approx 0.74$, so $u_U \approx (0.74 - 0.38)/(0.74 - 0.33) \approx 0.88$. Any u < 0.88 does not give U an incentive to strike. The P component of Eq. (7) can be solved for

$$u_P = (x' - C_1') / (C_2' - C_1'), \qquad (12)$$

which is the smallest u that prevents P incentives. At m = 5 , $u_P \approx (0.38 - 0.74)/(0.74 - 0.33) \approx 0.12$, the complement of u_U. This gives a wide band of probabilities that eliminate strike incentives in the interaction between U and P. If the first strike probability fell below 0.12 or rose above 0.88, the averaging at node 7 would produces expected costs smaller than those of inaction, and exchanges could result.

V. Stability.

The side that could strike first would see an incentive to increase its forces to levels where reduced costs produced an incentive to strike. That introduces arms control stability into what is otherwise a static discussion of first strike stability. In this metric the two types of stability are not cleanly separated. The knowledge that the other side might strike first restrains escalation and action. If either was unable or unwilling to strike first, its opponent would see an incentive to escalate, as that would decrease its costs, although it would also increase the probability of interaction. It is not enough to prefer inaction; each must assure that the other also prefers it.

Decisions at each node depend on strike costs and damage objectives, so each side is more likely to compete when its damage objective is large and it is likely to strike first. Those parameters dominate details of force composition, first and second strikes, and strike costs in the analysis above. However, they are difficult to quantify and not subject to direct observation, as one's decisions are largely determined by the opponent's objectives and strike probability. For example, at node 5, U's decision to strike depends on its assessment of whether P would strike, given the opportunity to do so. That depends on P's first strike cost, which U might be able to estimate accurately, and on P's damage objective, which U can only infer. Thus, each side's assessment of stability is determined by its estimate of the other side's damage objective and strike probability, which it does not know. Ignorance of those parameters has more impact than details of forces. Better ways of inferring an opponent's objectives and strike probability from history and events could significantly improve this situation.

Metrics. According to the cost ratio metric C_1/C_2 of Fig. 4, stability falls sharply for $m > 1$; each side evaluates stability by its analysis of its own forces, costs, and objectives; and both see incentives for escalation to large m. According to Fig. 7, game theory does not produce any incentive to strike below $m = 3.5$; both sides react to the other's poorly known objectives and decisions; and instability onsets abruptly when one side's first strike cost drops below its damage objective. Large m is a region of instability for both metrics, but game theory shows that both sides have incentives to avoid it.

The fundamental difference between cost ratio analysis and game theory is not strikes or costs; it is their decision variables. Cost ratios involve second strike costs in a fundamental way; game theory does not. Game theory is formulated in terms of first strike costs and damage objectives, which are the quantities on which each side must base first strike decisions. One does not plan to strike second; that is a by-product of the other's decisions. Since second strike costs can neither be used nor enforced, they are not appropriate decision variables. A variable that does provide a quantitative measure of stability is the difference between first strike cost and damage objective. It is positive when there is no strike incentive, negative when there is one, and zero at the transition; thus, $C_1 - x$ has the attributes of a proper measure of stability, and it generalizes naturally to more complex forces and situations. It is the fundamental decision variable in the game theoretic analyses above and below.

Analysis. Strategic games have a long history. Schelling proposed the form described above,[7] but a recent summary by Powell seems to cast doubt on their relevance in stating "Thus, the conventional logic of crisis stability implies that because there are first-strike advantages in this situation, there should be some risk of preemption. But as will be seen, there is no risk. The game is completely stable, even though there is an advantage to striking first."[8] That conclusion actually follows not from the intrinsic structure of the game, but from the payoffs assumed. Powell assumes that the payoff to losing is always preferable to that for striking first, which is equivalent to assuming C_1 is always larger than x. Under that condition the analysis above is also stable. However, Figs. 5-7 show that the costs produced by limited strikes on military value do not always satisfy that assumption.

Powell justifies the game theorists' assumption by stating that "...a general nuclear war would be so horrible that if a state had to choose between the certainty of war, even a war in which it was sure of having the first strike, and the certainty of peace, albeit a peace secured by submitting to its adversary, then the state would choose to avoid a general nuclear war." That is, a state would always submit rather than strike first, no matter what the cost. If the only engagements considered were general wars to the extinction of both sides, the assumption that $C_1 > x$ would be justifiable. However, the results above show that for current threats, which are largely confined to missiles and military recovery value targets rather than all-out attacks on populace, that is not the case. They involve costs commensurate with those of the military, economic, and political objectives held at risk. Because military analyses do not treat only general nuclear war, they do not generate the extreme costs assumed by academics, so they can lead to first strike costs below damage objectives.

Powell recognizes this possibility that in stating "The game can also be made to conform to the classical logic of war by assuming, for example, that the payoff to a successful first strike is greater than the payoff to submitting [i.e., $C_1 < x$]. Thus, the nuclear revolution can be parameterized by selecting different values for the payoffs, while the underlying structure of the game remains the same. One can therefore trace the effects of the nuclear revolution on the game's equilibria and on the dynamics of escalation by varying the game's payoffs."[9] These statements anticipate the extension to $C_1 < x$ carried out above and the legitimacy of using military strike costs for payoffs.

Powell further states "Expanding the scope of the analysis by parameterizing the nuclear revolution would mean that historical cases from before the nuclear revolution could be brought to bear to evaluate and refine the models. And if, in the course of that analysis, the models seemed to accord well with the cases antedating the nuclear revolution, then one might have more confidence in their ability to explain nuclear crises." Given the sensitivity of the above results to damage objectives that have not been and cannot be tested in the nuclear arena, correspondence to non-nuclear engagements could provide important certification of analyses. While the strike costs used above are inferred from nuclear strikes, they can also be interpreted in terms of the effects of massive conventional strikes.[10] Thus, the analysis above also provides an appropriate framework for the discussion of historical interactions.[11]

Summary. Game theory provides insight into the incentives for strikes, the role of preemption, and willingness to strike first. It provides measurable, controllable decision variables for deciding whether to strike or change forces. It illustrates the origin of crises in the attempts of independent opponents to avoid being struck first, which can force them to strike. Imperfect information aggravates the problem. External sources produce crises, but subsequent internally driven events can generate costs larger than those of the external sources. Game theory illustrates the role of uncertainty in reducing strike incentives by admixing second strike costs to offset the first strike costs that are the source of the incentives. The formalism is appropriate for strikes on military value, given appropriate costs, and that the analysis is the appropriate extension of academic game theory to non-nuclear engagements and the interpretation of historical information.

VI. Strategic Force Reductions

The impact of strategic force reductions on stability can be studied[12] by replacing the nominal U and P forces used above with those of the U.S. and Russia under SALT, START, and further reductions postulated by the National Academy of Sciences (NAS).[13] Table II shows how those forces are converted into the average numbers of weapons and launchers used in the exchange model in the Appendix. Although U.S., Soviet, and Russian force levels are used, it is not appropriate to identify forces with specific countries, as their damage objectives are not known and are as important as the forces. Thus, the two sides are identified only as U and P and assigned damage objectives x = x' = 0.38. The two strategic forces are broadly similar, although current U.S. forces have a greater fraction of survivable missiles. The large fraction of Russian submarines that are currently in port is a particular concern, as that converts survivable missiles and weapons into vulnerable targets that invite attack or preemption. The calculations below assume that 40% of U's submarines and 10% of P's are at sea at any time and that those in port do not participate in strikes.

Strikes and costs. Figure 8 shows the two sides' optimal allocations to vulnerable missiles, f and f'.[14] In moving from SALT to NAS, strategic forces fall by about an order of magnitude, while allocations to vulnerable missiles fall by factors of three, which reduces the weapons on vulnerable targets by an order of magnitude. U's allocation to missiles is about 50% larger in SALT through START II. P's numerous missiles permit it to allocate a larger fraction of its weapons to military value targets, but they represent an attractive nuisance to U. By START II(2) [i.e., START II, step 2], the two sides' allocations are approximately equal at f ≈ f' ≈ 0.3. That delivers 2-3 weapons on each vulnerable missile, whose survival probabilities fall to a fraction of a percent. Thus, vulnerable missiles are largely a sink, and second strikes are largely by survivable forces. By START III, allocations drop to about 0.1, which produces about one weapon per missile, so ≈ 10% of vulnerable missiles survive by leakage, although there are few vulnerable missiles left in post-START forces.

Figure 9 shows U and P's first and second strikes on value. The top curve is P's first strike; the second is U's first strike; the third is U's second strike; and the bottom is P's second strike. F' falls roughly linearly with period. F is smaller and roughly constant in SALT II through START I (2), after which it falls in parallel with F'. S is about half as large as F throughout due to U's large proportion of survivable forces. S' is small throughout due to P's submarine alert rate, which is 25% of U's; thus, S >> S' through NAS 2.

Figure 10 shows U and P's first and second strike costs. The top curve is P's second strike cost; the second curve is U's second strike cost; the third is P's first strike cost; and the bottom is U's first strike cost. U and P's second strike costs are similar in shape, although U's are lower due to its survivable force. The first strike cost curves differ by factors of 2 to 3 in early periods, but P's first strike costs start higher and slope downward, while U's starts lower and slopes upward, so the discrepancy narrows to about 50% by START II(2). P's first strike cost C_1' falls below its damage objective x' ≈ 0.38 at START II(2). U's first strike costs starts below x at SALT II, so $C_1 < x$ in all periods shown, which has implications for strike incentives discussed below.

Graph. Figure 11 shows the cost of nodes 1 and 2 in each period. Node 1 is a first strike by U followed by a second strike by P with costs (C_1, C_2'). P does not strike at node 2 through START II(1) because Fig. 10 shows that C_1' > x', so costs are (x, x') there. In START II(2), C_1' falls below x', so P has an incentive to strike, which produces a spike. In START III, C_1' > x', and the incentive disappears. In NAS 1 & 2, C_1' < x', and it reappears. Figure 10 shows that P's cost reduction C_1' - x' would be small for any of the incentives. They are small and represent differences between large, uncertain quantities, but they illustrate the sensitivity of decisions to costs and objectives that are not known precisely.

If P could strike first at node 2, U's cost would jump from x to C_2, the second curve on Fig. 10, i.e., it would increase by $C_2 - x \approx 0.73 - 0.38 \approx 0.35$, roughly doubling it. Thus, a decision by P on the basis of small incremental costs could greatly increase U's cost. However, on the upper branch, only U can strike first. Recognizing that P will strike at node 2, given the option, U preempts at node 5. Because $C_1 < x$ in every period, U would strike first in each. That increases P's cost to C_2', which roughly doubles them. Figure 12 shows the resulting costs at node 5, which are (C_1, C_2') in every period.

On the lower branch the roles of U and P are interchanged, so P can strike first. The costs of nodes 3 and 4 are shown in Fig. 13, which is the same as Fig. 10 except for re-labeling. Node 3 is a strike by P and a restrike by U, which has costs (C_2, C_1'), the second and third curves down. At node 4, U decides whether to strike by comparing C_1 with x. Since $C_1 < x$ in all periods, U would always strike first, if given the option, which would produce costs (C_1, C_2'), the top and bottom curves. However, on the lower branch, U cannot strike first. Recognizing that U will strike first if allowed to, P strikes first at node 6, producing costs (C_2, C_1') in every period, the first and third curves on Fig. 13. In periods before START II(2), C_1' > x', so P would not see an incentive to strike, if the alternative was inaction. However, if P did not strike, U would, so P is forced to strike first in every period, even though the cost of doing so, C_1', is greater than the cost of inaction, x', because the alternative of not striking has a cost C_2'>> C_1'.

U and P's costs at nodes 5 and 6 in each period are summarized in Fig. 14, which is again a re-labeling of the first and second strikes of Fig. 10. The two middle curves are from P's decision to preempt at node 6 on Fig. 13; the two exterior ones are from U's decision to strike at node 5 on Fig. 12. P has an incentive to strike when $C_1' < x'$, and U has an incentive when $C_1 < x$. The former holds after START II(1); the latter holds throughout. If P was certain it could strike first, it would in all periods after START II(1). U would have an incentive to strike first in all periods for the damage objectives assumed.

Whether U or P can strike first is decided by Nature. Figure 15 shows U and P's expected costs at node 7 for u = 0.5, i.e., both sides equally likely to strike first. The top curve is P's cost, 7P,which is greater than x' = 0.38 in all periods, so P has no incentive to strike. The bottom curve is U's expected cost, 7U, which is also greater than x in all periods. Averaging U's very low first strike cost with its appreciable second strike cost with u = 0.5 increases 7U enough to remove the incentives that would exist in all periods if U was certain it could strike first. However, 7U - x is large only for START II(2). In other periods it is < 10%, and it falls in going from NAS 1 to NAS 2, which suggests that incentives could reappear at lower force levels. Investigating that would require assessment of damage objectives at those levels, which could differ from those at larger force levels.

Breakeven strike probability. Maintaining 7U > x to avoid incentives places limits on u. The range of allowable values of u can be determined by inverting the equation for the expected costs at node 7. The U component gives

$$u = (7U - 6U)/(5U - 6U). \qquad (13)$$

Substituting x for 7U produces the "breakeven" u_U, which is the maximum value for which U would see not see an incentive to strike, the top curve in Fig. 16. It gives $u_U \approx 0.6$ for periods other than START II(2), where it increases to ≈ 0.7. The P component gives

$$u = (7P - 6P)/(5P - 6P). \qquad (14)$$

Substituting x' for 7P gives u_P, the minimum value for which P would not see an incentive, the bottom curve on Fig. 16. It increases almost monotonically with period, reaching ≈ 0.05 to 0.15 in NAS 1 & 2. It is negative in periods earlier than START II(2) where P's costs are so high that it would take an unphysical, negative value of u to reduce expected costs to levels where strikes would appear appropriate. However, in NAS 1 & 2, P's first strike costs are small enough for it to see an incentive to engage a U that was unlikely to strike first.

The reason u_P increases in later periods is shown by the first strike costs of Fig. 10. After START II, 5P = C_2'and 6P = C_1', so $u_P = (x' - C_1')/(C_2' - C_1')$. The numerator and C_1' are about the same for START II(2) and NAS 1 & 2, so the variation of u_P is determined by the denominator. Figure 10 shows that C_2' decreases by about a factor of two over this interval, which gives the corresponding increase in u_P. C_1' is roughly constant because as its damage to self falls in later periods because U has fewer weapons, P's cost due to incomplete damage to U increases as P runs out of weapons. C_2' falls because P can still achieve its desired damage to U, but the damage U can do to P decreases rapidly as it runs out of weapons. Thus, u_P must increase in later periods to compensate for U's reduced ability to do damage with the reduced number of weapons it has there.

The band of probabilities shown in Fig. 16 are for baseline damage objectives (L, L') = (0.6, 0.6); they vary for other values. For (L, L') = (0.3, 0.6), i.e., a less aggressive U, u_U approaches unity, while u_P has roughly the same shape as in Fig. 16. For (L, L') = (0.6, 0.3), a less aggressive P, u_U is as in Fig. 16, but u_P is negative in all periods, i.e, U does not have to be willing to strike first to achieve stability. For (L, L') = (0.3, 0.3), u_P is negative in all periods, and u_U approaches unity in periods other than START II(2) (which requires $u_U = u_P = 0$); thus, a factor of two reduction in both sides' damage preferences shifts from a region where there are significant constraints on U's strike probabilities to one that is stable for any willingness to strike first.

Summary. Large forces produce decisions that differ somewhat than those in modest, symmetric forces. The differences are primarily due to asymmetries in the survivability of current strategic forces. Vulnerable forces act as an attractive nuisance to the more survivable side; they offer a temptation to suppress damage while accumulating value. For current forces, the more survivable side has a first strike cost below its damage objective throughout, so it has an incentive to strike first whenever it can. The vulnerable side only sees incentives to strike in later periods, but is forced to strike whenever it can by the survivable side's incentive to strike if it does not. The net result is that whichever side can strike first in a crisis does so, which burdens their interaction with strike incentives to both sides.

These incentives are removed by averaging over the uncertainty of which side might strike first in a crisis, which averages high second strike costs with the low first strike costs that produce the incentives. However, that generally requires that both sides be willing to strike in crises. The two sides' first and second strike costs determine the limits on the probability of striking that are necessary to eliminate incentives for both sides. Nominal damage objectives place significant limits on that probability, particularly in later stages of arms reductions. Factor of two reductions in damage objectives could eliminate those restrictions and possibly the necessity to strike first in a crisis. Conversely, reduction of second strike costs in later periods makes averaging less effective, which means that lower levels of less vulnerable strategic forces are not necessarily more stable.

VII. Trilateral Offensive Forces

Trilateral conflicts force each side to divide its forces between simultaneous bilateral engagements.[15] The configuration analyzed below is motivated by current concerns about the increase of Chinese missile forces in an environment in which China and Russia remain independent but see strong economic and political reasons to avoid conflict. In it, the trilateral strategic balance is dominated by two bilateral balances—one between the U.S. and Russia and another between the U.S. and China—which force the U.S. to divide its forces between the two. As the three sides' objectives are not known, they are identified only as U, P, and T. Side U and P's forces correspond to U.S. and Russia's ≈ 5,000 weapon forces of START I(3). Side T's forces follow a reverse progression from NAS 2 level back through the earlier steps of START, i.e., they progress from 300 weapons up to levels comparable with those of U and P. As objectives are not known, it is not appropriate to identify the sides with specific countries.

Trilateral exchanges are modeled and optimized[16] with extensions of the exchange models, strike optimizations, and conversion of damages into first and second strike costs and indices derived earlier and described above.[17] Side i, where i = 1-3, has M_i = vulnerable missiles with m_i warheads and N_i survivable missiles with n_i warheads each for a total of $W_i = m_iM_i + n_iN_i$ warheads. It is assumed that all weapons can be used in a first strike of magnitude W_i, but only a fraction h_i of the N_i survivable missiles can be used in a second strike, i.e., that a fraction $1 - h_i$ of them are vulnerable when off alert and cannot be launched on warning or under attack . In striking first, side i uses a fraction g_{ij} of its missiles in the attack on side j, where $\Sigma_j\, g_{ij} = 1$ for each i. In the case where P and T strike only U, $g_{up} = 1 - g_{ut}$, and $g_{pu} = g_{tu} = 1$, as no side values residual weapons. Side i delivers a fraction f_{ij} of its weapons on j's vulnerable missiles and the remaining fraction$1 - f_{ij}$ in a first strike on j's military value targets of magnitude

$$F_{ij} = (1 - f_{ij})g_{ij}(m_iM_i + n_iN_i). \tag{15}$$

The average counter force strike of $r_{ij} = f_{ij}g_{ij}W_i/M_j$ warheads on each j vulnerable missile produces a survival probability $Q_{ij} \approx q\hat{}(r_{ij})$, where $p_{ij} = 1 - q_{ij}$ is i's single shot kill probability against a j missile (taken to be 0.8 for all missiles). Side j's second strike on i is

$$S_{ji} = Q_{ij}m_jM_j + h_{ji}n_jN_j, \tag{16}$$

which is delivered on value, as residual weapons have no value. Side j's first strikes F_{ji} and i's second strikes S_{ij} are obtained by conjugating Eqns. (15) and (16). First and second strikes are converted into costs through exponential approximations to the value of military value targets assuming $1/k_i \approx 1{,}000$ for all i. The cost of damage to self and other are joined with weighting parameters L_{ij} that measure the attacker's relative damage preference. The calculations below use $L_{ij} = 0.6$, (i.e., $x_{ij} = 0.38$) for i = U and j = P or T and i = P or T and j = U, and $L_{ij} = 0$ otherwise. The cost to side i for striking first if j strikes back is

$$C_{1ij} = (1 - e^{-kiSji} + L_{ij}e^{-kjFij}) / (1 + L_{ij}). \tag{17}$$

If there is no interaction, i's first strike costs reduce to

$$C_{1ij} = L_{ij} / (1 + L_{ij}) = x_{ij}, \tag{18}$$

the cost to i for not fulfilling its damage objective with respect to j or damage objective . The cost to i for striking back after j strikes is

$$C_{2ij} = (1 - e^{-kiFji} + L_{ij}e^{-kjSij}) / (1 + L_{ij}), \tag{19}$$

which also reduces to x_{ij}, if neither side strikes. This is a direct extension of the bi-lateral formalism to tri-lateral engagements. The metric is the optimal use of U's forces in each engagement.

Strikes. The division of each sides' forces between engagements and their allocation between missiles and value targets in each are shown in Fig. 17. When T's forces are small, U uses about 55% of its forces on P. As T's forces approach U and P's, the fraction of U's forces used against P drops to about 50%. Of the forces U directs to P, about 45% are allocated to P's vulnerable missiles, as in U and P's bi-lateral interaction. P's allocation to missiles remains at its bi-lateral value of ≈ 35% throughout. U increases the fraction of its forces used against T from about 45% at NAS 2 to 50% at START I(2). U's allocation to T missiles is about 5% at NAS 2; it increases with T's forces to parity with its 40% allocation to P when T reaches START I(2). T's allocation to U missiles is zero for NAS 1 & 2, where T cannot damage limit effectively. It is comparable to P's by START II(2) forces levels. The fraction of U's forces used on T is large even when T's forces are small because T's target set is assumed to be as large as P's. The resulting divisions maintain the separation of the engagement between U and P and that between U and T.

Figure 18 shows the three sides' first strikes on value as T's forces vary from NAS 2 to SALT levels. U and P's first strikes on each other are relatively constant, as their total forces do not vary and the fractions used on each other vary little. U's first strike on T is about 2,200 weapons at NAS 2, where T's forces are small and U's damage limiting is effective; it falls to parity with U's strike on P at START I(3), where U, P, and T's forces are comparable. T's first strike on U increases monotonically with its forces. At NAS 2 it is all 300 of T's weapons. At START I(3), it is comparable to P's 3,400 weapon first strike on U, so P and T's first strikes are about equal. U must divide its forces between the other two, so it delivers about 1,600 weapons on each.

Figure 19 shows their second strikes. P's restrike on U is $S_{pu} \approx 300$ weapons, which is constant and small due to U's suppression of P's vulnerable missiles—particularly those off alert. U's second strike on P is $S_{up} \approx 1{,}500$ weapons, which is larger because of U's survivable weapons. U's restrike on T is ≈ 3,000 weapons at NAS 2; it falls to ≈ 1,500, the level of U's restrike on P, when T reaches START I(3). T's second strike on U is less than 100 weapons at NAS 2; it approaches U's second strike at START I. It approaches U's second strike rather than P's due to the assumption that T's alert rate for survivable forces is the same as U's rather than P's. If T's alert rate was P's 10%, T's second strike would approach P's at START I.

Costs. Figure 20 shows first strike costs. U and P's are roughly constant because their forces and allocations vary little. The cost to U of striking P first is $C_{1up} \approx 0.23$, which is less than U's damage objective of with respect to P, $x_{up} \approx 0.38$, because of P's vulnerable forces. P's first strike cost on U is $C_{1pu} \approx 0.5$, which is larger due to U's survivable forces. The cost of U preempting T is C_{1ut}, which is ≈ 0.1 when T is at NAS 2 levels. It increases to ≈ 0.5 when T reaches START I(3) because T's restrike is then comparable to P's and additive to it. The cost of T preempting U is ≈ 0.9 at NAS 2; it falls to the ≈ 0.5 cost of U striking T and P striking each other START I(3). U's first strike cost is below its damage objective of ≈ 0.38 with respect to T after START II, which produces strike incentives that are discussed below.

Graph. This three-sided interaction dominated by two bilateral interactions can be described by two graphs, one for the interaction of U and P and another for the interaction of U and T, as P and T are assumed not to interact. The principal coupling between the two conflicts is U's division of its forces between the two conflicts, which is determined by assuming a trial division, minimizing each side's first strike cost in each graph separately, and iterating the division to determine the value that minimizes first strike costs in each. Because forces and allocations vary little, the graph for U and P is similar to that for their bilateral interaction at these force levels in the previous section, as are the expected costs at node 7 in Fig. 21. $C_{1up} < x_{up}$ throughout, so U always sees an incentive to strike P, but averaging with a first strike probability $u = 0.3$, i.e, P twice as likely to strike first, produces U and P expected costs of ≈ 0.55 to 0.58 throughout, which are well above their costs of inaction, $x_{up} = x_{pu} = 0.38$. Neither side sees an incentive to strike, so it is not necessary to discuss the graph for U and P in detail.

In the interaction between U and T, the dominant consideration is that $C_{1ut} < x_{ut} = 0.38$ for START III, NAS 1, and NAS 2 level T forces, so U has an incentive to strike in those periods. Figure 22 shows the cost of nodes 1 and 2. Node 1 is a strike by U and restrike by T with costs (C_{1ut}, C_{2tu}). Node 2 is a strike by T and restrike by U; however, $C_{1tu} > x_{tu}$, so T does not strike, and the costs are (x_{ut}, x_{tu}). At node 5, U minimizes its cost by choosing the minimum of C_{1ut} and x_{ut}. That is C_{1ut} for START III through NAS 2, which produces the costs in Fig. 23. By striking, U decreases its cost by $x_{ut} - C_{1ut} \approx 0.1$ to 0.3 but increases T's by $C_{2tu} - x_{tu} \approx 0.4$ to 0.5. Thus, a small cost reduction to U roughly doubles T's cost. Moreover, U's low first strike cost for small T forces gives it an incentive to strike before they can grow to levels where damage limitation loses effectiveness.

On the lower branch U and T 's roles are interchanged. Node 3 is a first strike by T and restrike by U, which produces costs (C_{2ut}, C_{1tu}). At node 4, U decides whether to strike by comparing the cost of striking C_{1ut} with that of inaction, x_{ut}. As C_{1ut} is the smaller for T forces smaller than START II, U would strike there, producing node 4 costs (C_{1ut}, C_{2tu}). However, on the lower branch, U does not have the option to strike first. At node 6, T chooses the lesser of the costs of inaction, C_{2tu}, the top curve of Fig. 24, and that of striking first at node 6, C_{1tu}, the curve below it. Since striking first has the lower cost, and U will strike if it does not, T strikes at node 6, producing costs (C_{2ut}, C_{1tu}) in NAS 2 through START III, as shown in Fig. 25.

For force levels below START III, T strikes even though the cost of doing so, C_{1tu}, is larger than that of inaction, x_{tu}, because not striking would incur the still greater cost $C_{2tu} >> C_{1tu}$. T's preemption to avoid a strike by U produces much higher cost than that of inaction, and by subjecting U to a first strike, it also inflicts a much higher than inaction on U. U's incentive to strike is the cause of action, but on the lower branch, it acts through the incentive it gives T to preempt.

Whether U or T can strike first is decided by Nature. Figure 26 shows U and T's expected costs at node 7 for $u_{ut} = 0.2$, i.e., T four times as likely to strike as U. The parallel lines are U and P's expected costs, which are well above those of inaction, indicating that they see no incentive to engage. That remains so even when half U's forces are diverted to T, which indicates the robustness of this result. The top curve is T's cost, 7T; the bottom is U's cost, 7U. For large forces their average is (x_{ut}, x_{tu}). For small forces it is $u_{ut}(C_{1ut}, C_{2tu}) + (1 - u_{ut})\ (C_{2ut}, C_{1tu})$. Because C_{2tu} and C_{1tu} are both ≈ 0.7 to 0.9 and roughly equal beyond START II, T's expected costs there at node 7 are ≈ 0.7 to 0.9, which are much larger than those of inaction. T sees no incentive to engage for any u_{ut} because U's incentive to strike at node 5 cannot be preempted. While U's incentive to strike at node 4 gives T an incentive to preempt at node 6, doing so would involve costs so great that T would prefer not to engage at all.

U's costs at node 5 decrease rapidly after START II, and those at node 6 increase only slightly in START II and NAS 1. After averaging, U's expected costs at node 7 are greater than inaction for START III and NAS 1, but less than those of inaction at NAS 2 and later, where U sees an incentive. Its magnitude depends on the specific forces and parameters used. At NAS 1, $C_{1ut} \approx 0.1$ and $C_{2ut} \approx 0.2$ are both < 0.38, so U's expected cost at node 7 is $< x_{ut}$ for any u_{ut}; thus, U always sees an incentive to strike. Both U and T's incentives stem from $C_{1ut} < x_{ut}$. Reversing that inequality would require that U's first strike cost be increased or that its damage objectives be decreased. Further increasing C_{1ut} at small T would be difficult, as the forces assumed already have significant survivability. Reducing U's damage objective with respect to T is possible. Fig. 20 shows that a value of x_{ut} less than 0.1 would remove U's incentive to strike at nodes 2 and 4 through NAS 2. With that incentive removed, T would see no incentive to preempt at node 6, so there should be no interaction and costs should remain (x_{ut}, x_{tu}) in all periods.

The result that a strong country sees an incentive to strike another that is threatening while its forces are too small to mount a significant strike has been referred to as "killing the viper in its nest." The above results indicate that such an incentive occurs at small force levels and enters in two ways. On the upper branch it gives the stronger side an incentive to suppress the smaller; on the lower branch it causes the smaller side to preempt. This strike incentive at small forces produces two interactions, both of which are untoward. That such strikes have not occurred suggests that damage objectives are smaller than those assumed. Above, U's damage objectives are of greatest concern, as they produce the strike at nodes 5 and incentive to strike at node 4 that forces T to preempt at node 6. To the extent that U can be identified with the U.S., it should be possible to reduce or communicate its objectives sufficiently to eliminate incentives. Damage objectives below 0.1 would eliminate incentives for the forces used. At still lower force levels, damage objectives would have to be reduced further. For few and less survivable T weapons, the relationship between U and T would have to be non-aggressive.

This example is nominally trilateral, but it largely decomposes into the two interactions treated above. The first is the interaction between U and P, which resembles their bilateral engagement, even when U reallocates significant forces to T. It remains stable despite significant apparent strike incentives because of the averaging due to uncertainty discussed above. The second is the interaction between U and T, which resembles the bilateral interaction between two sides at START and NAS force levels. If T is willing to preempt, it can increase U's costs at intermediate force levels to levels that make inaction preferable, but for small T force levels, U sees an incentive to strike for the damage objectives postulated. As the interaction between U and P does not induce incentives, and that between U and T does, the stability of such a triad is governed by that of the least stable bilateral interaction in it. That contrasts with the result obtained if P and T are combined into a single side with total force P + T. Then, T makes a negligible contribution, stability is determined by the interaction between U and P + T, and the interaction would be stable for all forces shown.

This analysis can be used to discuss in an approximate fashion the impact of maintaining U.S. forces at START levels while Russian forces decrease significantly for economic or other reasons. Take U to be the U.S,. T to be Russia, and ignore P. Then, as Russia's forces (T's in the charts above) decrease, Fig. 20 shows that the U.S. would have an incentive to strike Russia once its forces fell below START II levels. Figure 25 shows that Russia would have an incentive to preempt at node 6. That would stabilize its transition through START III and NAS 1 level forces, but Fig. 25 shows that even a reluctant U.S. would have an incentive to strike Russian NAS 1-level forces, even if they were more survivable than at present. This interpretation is no more certain than knowledge of U.S. and Russian damage objectives and strike probabilities, which could be lower than those assumed. In general, when either side becomes less aggressive, it sees less incentive to engage.

The growth of a third power does not strongly perturb the interaction between the two major powers, whose interaction remains largely bilateral. However, for T small, U can generate a large first strike on T by diverting a small fraction of its forces, while T can only deliver a small second strike. Thus, U's first strike costs are small, which generates an incentive for it to strike T before its forces can grow. Recognizing that, T has an incentive to preempt even when the cost to it for doing so is very large. The problem is not the preemption of T by U, but T being forced to preempt U. U sees an incentive to preempt when it can, which forces T to preempt if it can. Their interaction generally involves strike incentives on both sides that are not removed by averaging over first strike probabilities. The only obvious way to eliminate them is to reduce U's damage objective, as T's damage objective, strike probability, and forces are not constraints in the analysis.

VIII. Summary and Conclusions

Game theory provides insight into the roles of strikes, preemption, uncertainty, and willingness to strike first, and defines relevant , measurable decision variables for their integration. It shows that crises are generated internally by the attempts of independent players, none of whom controls all decisions, to avoid the cost of being struck, which can give them incentives or force them to strike. Striking first can reduce or increase the first striker's cost, but it avoids the greater cost of being struck, which provides a strong incentive to strike when one can. Uncertainty about which side can strike first reduces these incentives by admixing second strike costs that offset the small first strike costs that produce incentives. The analysis permits representative costs for military strikes on value to be used as the payoffs in an extension of academic game analysis. The resulting framework should apply to non-nuclear applications and be relevant for analyzing historical data to infer opponents' damage objectives and strike probabilities.

In strategic force reductions, the main concern is the vulnerability of current forces, which acts as an attractive nuisance. That is mitigated by uncertainty about which side might strike first in a crisis. Plausible first strike probabilities produce modest incentives, although eliminating them generally requires that both sides be willing to strike in crises. Nominal damage objectives place limits on permissible first strike probabilities, but reduced damage objectives could eliminate them. In later periods averaging becomes less effective, which suggests that smaller, less vulnerable strategic forces are not necessarily more stable.

The growth of a third side does not strongly perturb the interaction between two strong powers, whose interaction is still largely bilateral. A strong side can generate a large first strike on the weak one by diverting a small fraction of its forces, while the weak one could only generate a small second strike with all of its forces. That produces incentives for the strong side to strike the weak before its forces can grow. Recognizing that, the weak side has an incentive to preempt even when the cost to it for doing so is very large. These dual incentives produce costs that deter strikes at intermediate force levels, but leave strong incentives at very small force levels.

At some level in the buildup of a weak force or the reduction of a strong one, the incentives produced by the weaker will produce strike incentives to the strong side, unless the weak side's costs or the strong side's damage incentives are reduced significantly. Further reducing the weak side's cost would be difficult. As the weak side's damage objective and strike probability are not constraints in the analysis, the only obvious way to reduce the strong side's incentive is to reduce its damage objective with respect to the weak side, which would require both reduction of its damage preference and communication of that reduction to the other sides.

References

[1] G. Canavan and J. Immele, “Stability Against Strategic Reconstitution by Transparency,” *Stability Roundtable II*, USSTRATCOM, July 1999; Los Alamos National Laboratory LA-UR-2636, August 1999.

[2] G. Canavan, “Crisis Stability and Strategic Defense” *Proceedings of the Military Modeling and Management Session of the ORSA/TIMS National Meeting, November 12–14*, S. Erickson, Ed. (Operations Research Society of America: Washington, 1991).

[3] G. Kent and R. DeValk, "Strategic Defenses and the Transition to Assured Survival," RAND Report R-3369-AF, October, 1986.

[4] R. Powell, *Nuclear Deterrence Theory* (Cambridge, University Press, 1990).

[5] T. Schelling, *The Strategy of Conflict* (Cambridge, Mass; Harvard University Press, 1960).

[6] G. Canavan and J. Immele, “Costs in Symmetric Strategic Games,” *Stability Roundtable II*, USSTRATCOM, July 1999; Los Alamos National Laboratory LA-UR-99-3997, August 1999.

[7] T. Schelling, *The Strategy of Conflict* (Cambridge, Mass; Harvard University Press, 1960).

[8] R. Powell, *Nuclear Deterrence Theory. op. cit.*, p. 113.

[9] R. Powell, *Nuclear Deterrence Theory. op. cit.*, p. 185.

[10] G. Canavan, “Stability of Nuclear and General Purpose Forces,” Los Alamos National Laboratory Report LA-UR-97-4618, November 1997.

[11] G. Canavan, “Stability of Nuclear Forces Versus Weapons of Mass Destruction,” Los Alamos National Laboratory Report LA-UR-97-4987, December 1997.

[12] G. Canavan, “Considerations in Missile Reductions and De-alerting,” L. Kruger ed, *Proceedings of the World Federation of Scientists Working Group on Missile Proliferation and Defense* (World Federation of Scientists, Lausanne, 1998); Los Alamos National Laboratory Report LA-UR-98-1426, April 1988.

[13] *The Future of U.S. Nuclear Weapons Policy*(Committee on International Security and Arms Control, National Academy of Sciences, 1997).

[14] G. Canavan, “Cost Variations in Strategic Force Reductions,” Los Alamos National Laboratory Report LA-UR-99-4579, August 1999.

[15] G. Canavan, “Cost Variations in Simultaneous Strategic Force Changes,” Los Alamos National Laboratory Report LA-UR-99-6483, January 2000.

[16] G. Canavan, " Considerations in Missile Reductions and De-alerting,” op. cit.

[17] G. Canavan, “Crisis Stability and Strategic Defense,” op. cit.

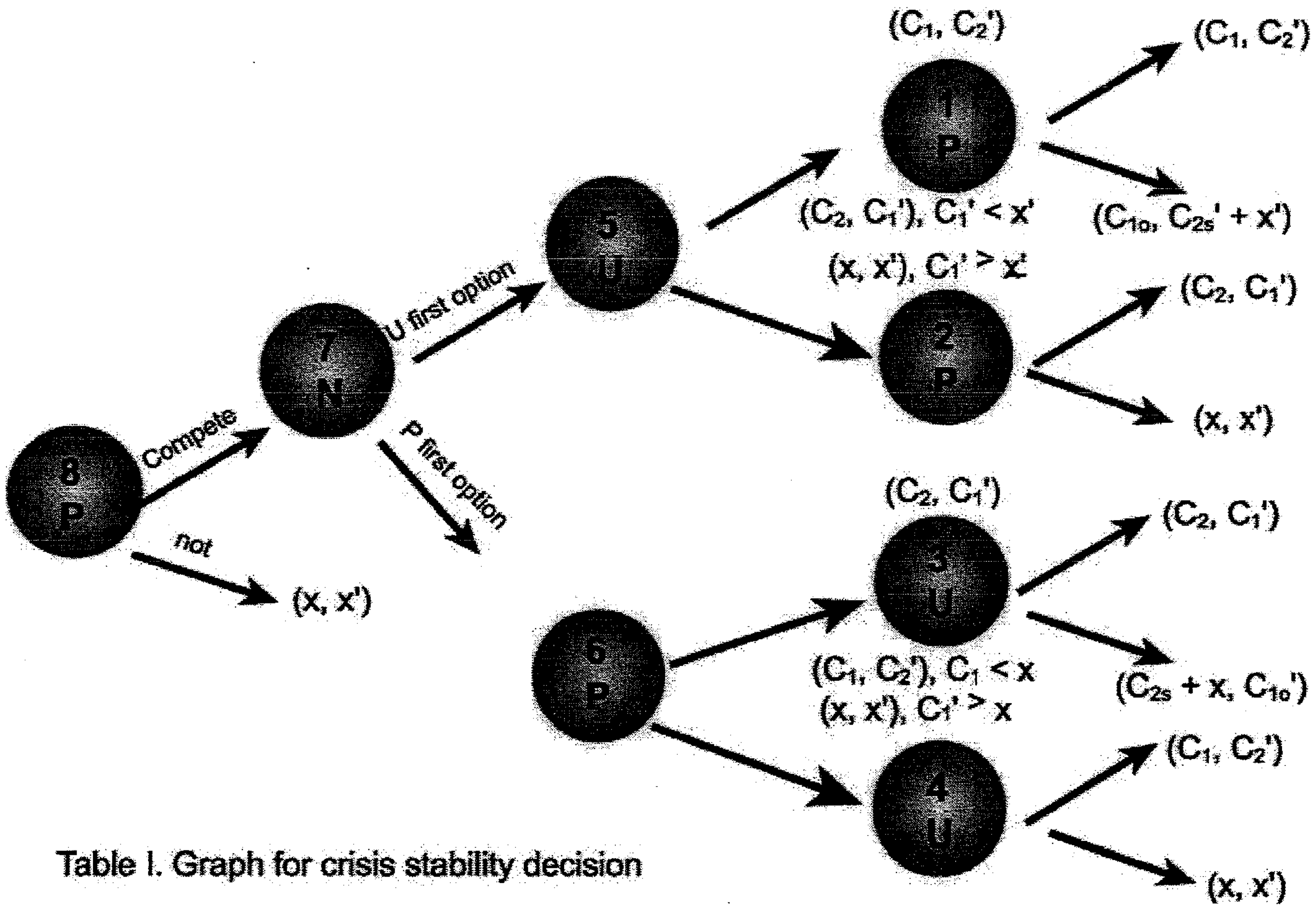


Table I. Graph for crisis stability decision

revised 2/28/01

Table II. Vulnerable and survivable missiles and weapons

pd.	regime	m	M	n	N	W	m'	M'	n'	N'	W'
1	SALT II	2.5	1000	8.6	672	8229	4.7	1400	3	940	9400
2	START I 1	3.2	960	8.2	528	7363	5.6	1090	3.5	728	8652
3	START I 2	3.2	755	8.1	480	6304	5.2	880	3.8	664	7099
4	START I 3	2	750	8	450	5100	3	880	3.8	660	5148
5	START II 1	1	500	8	375	3500	3	500	4	500	3500
6	START II 2	1	500	5	336	2180	1	500	5	450	2750
7	START III	1	200	4	336	1544	1	200	4	300	1400
8	NAS 1	1	160	3.5	240	1000	1	160	3.5	240	1000
9	NAS 2	1	60	1	240	300	1	60	1	240	300

Fig. 1. Allocation f, weapon/missile r, & surv prob Q

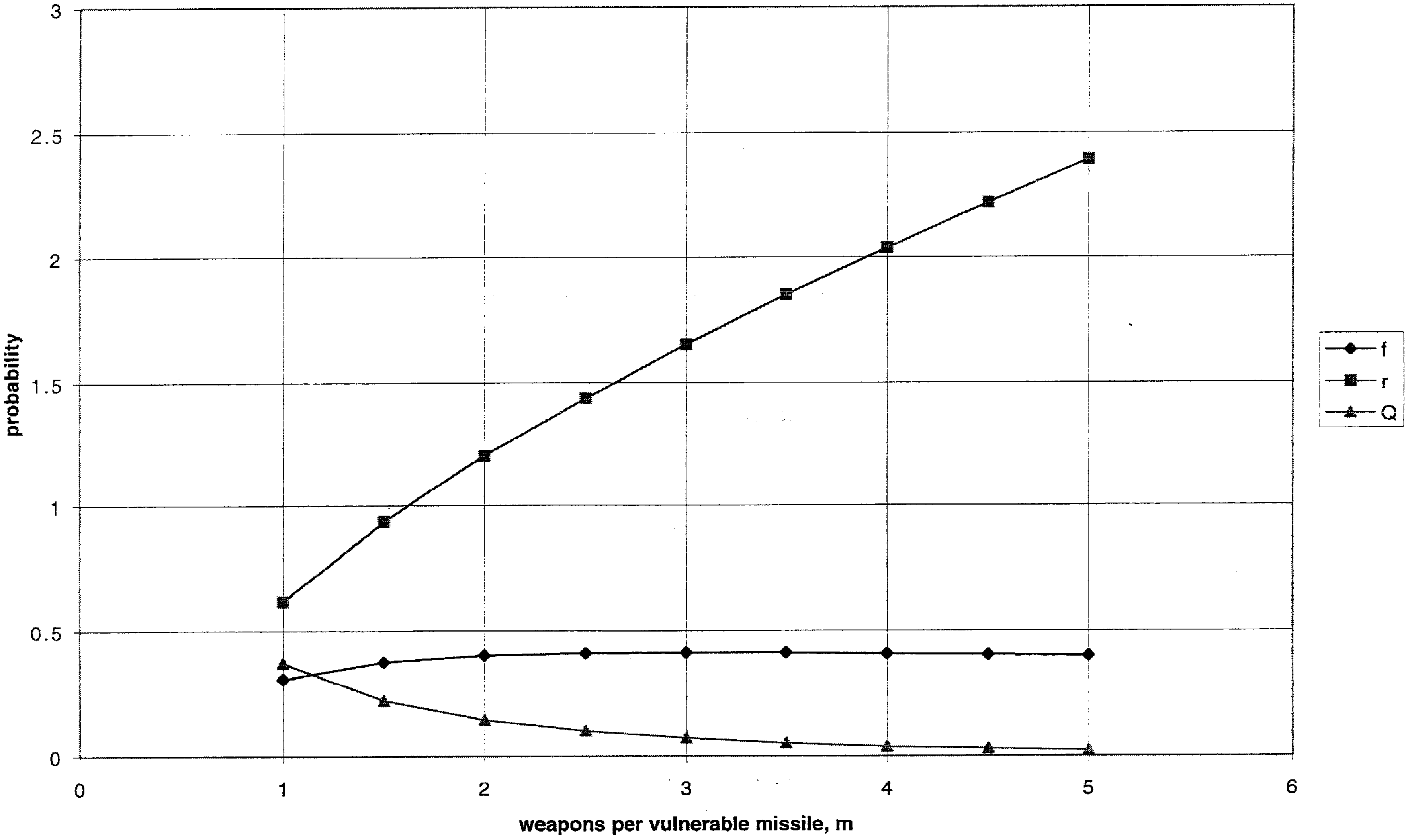


Fig. 2. First and second strikes

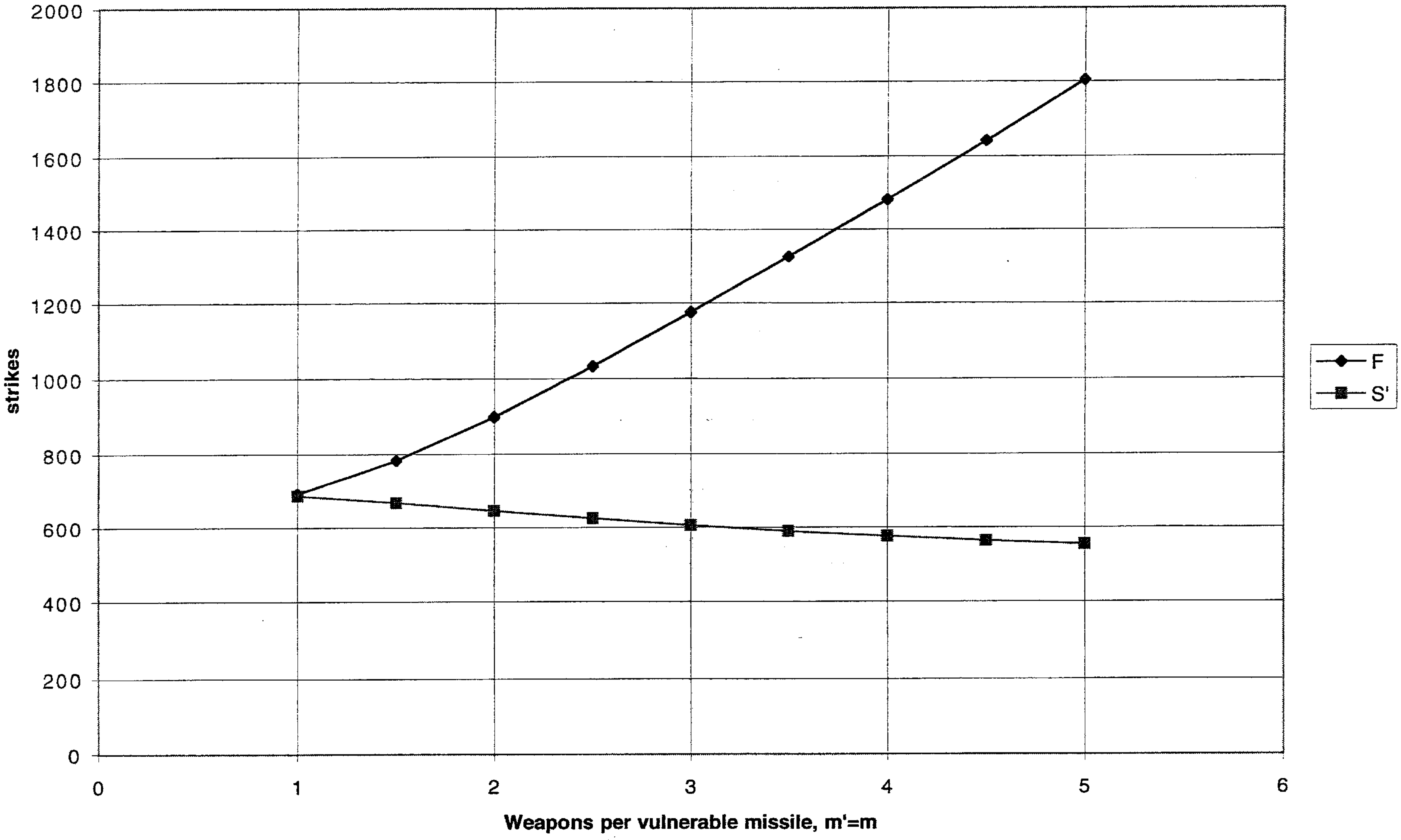


Fig. 3. First and second strike costs

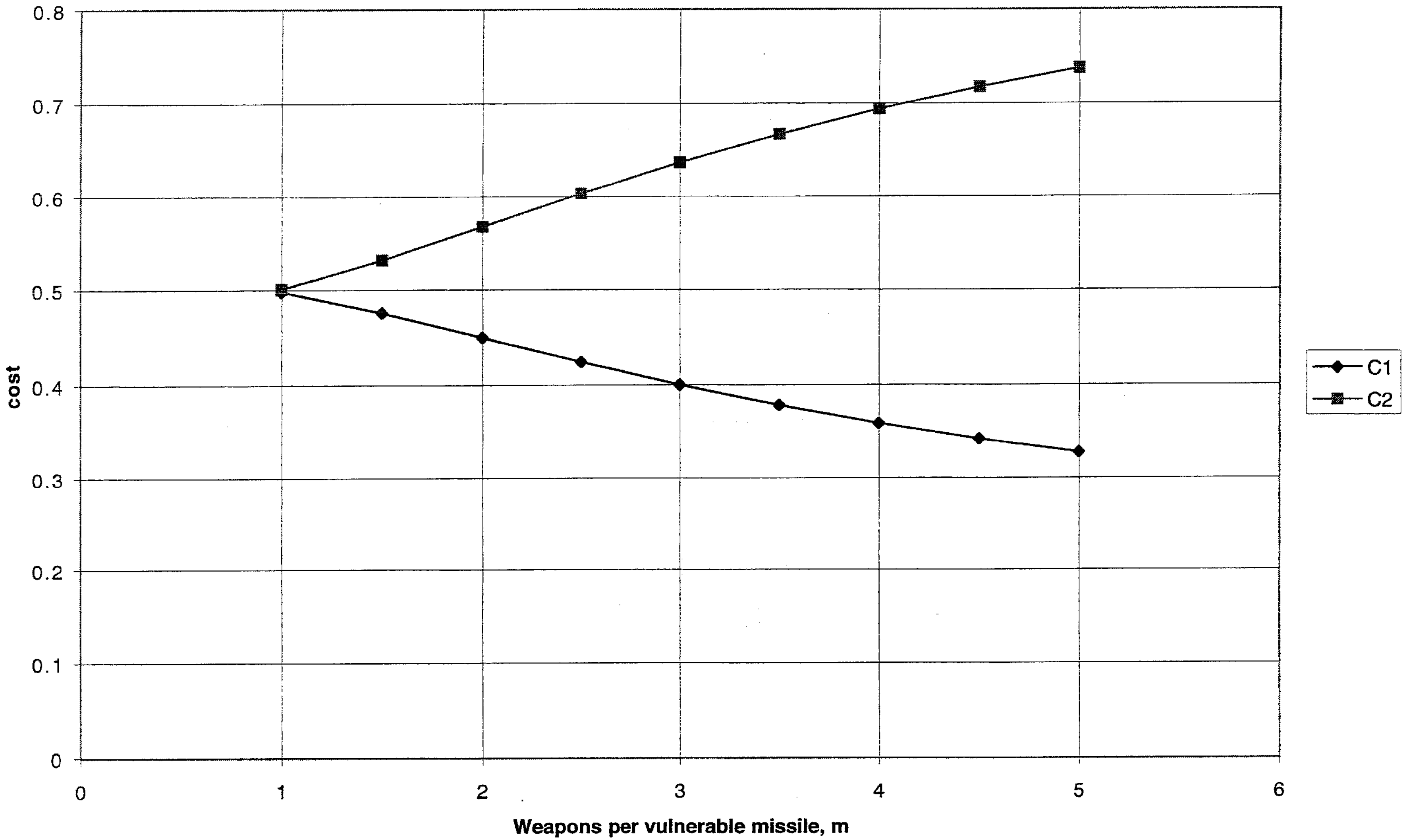


Fig. 4. Stability indices

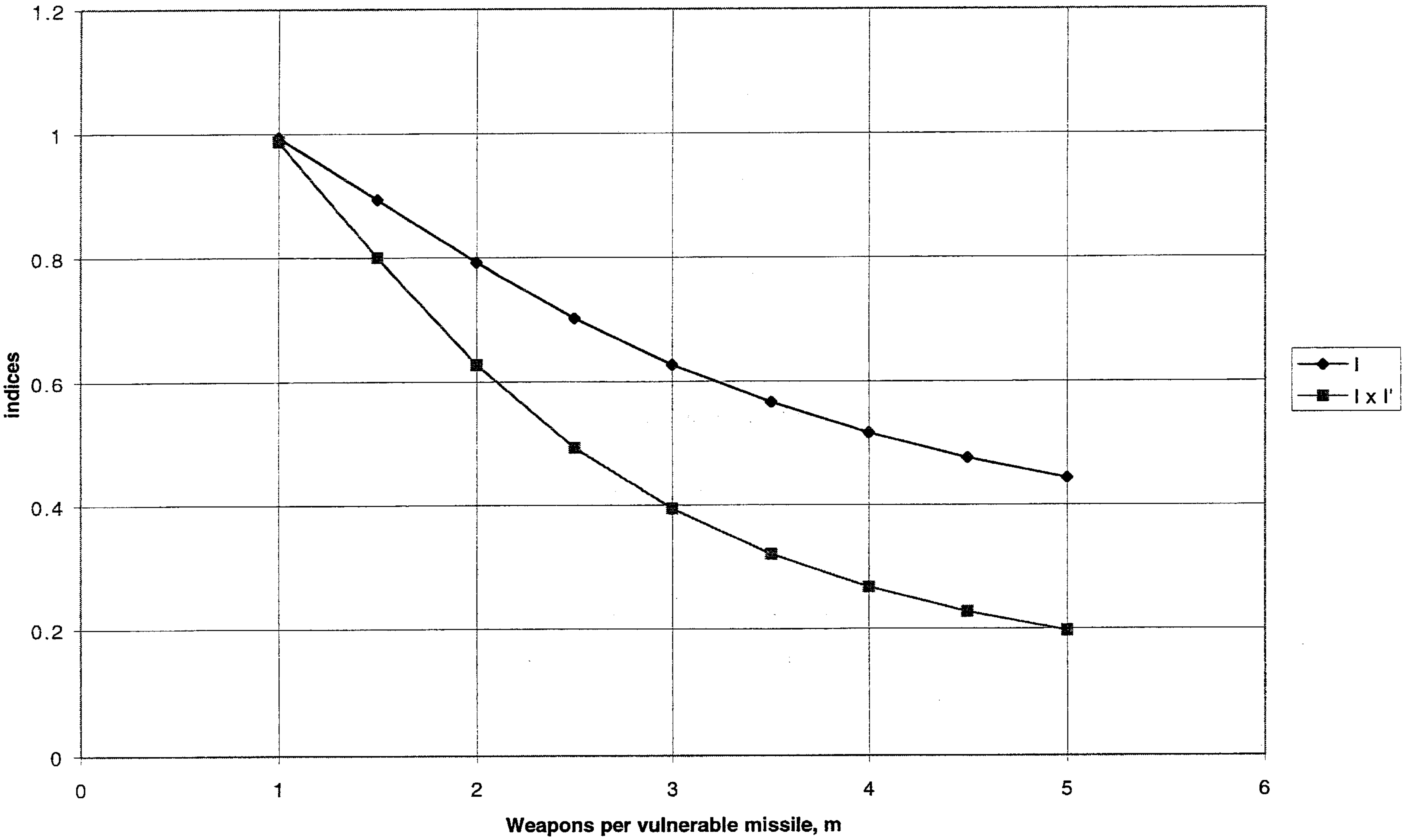


Fig. 5. cost of nodes 1 & 2 vs m' (U first strike option)

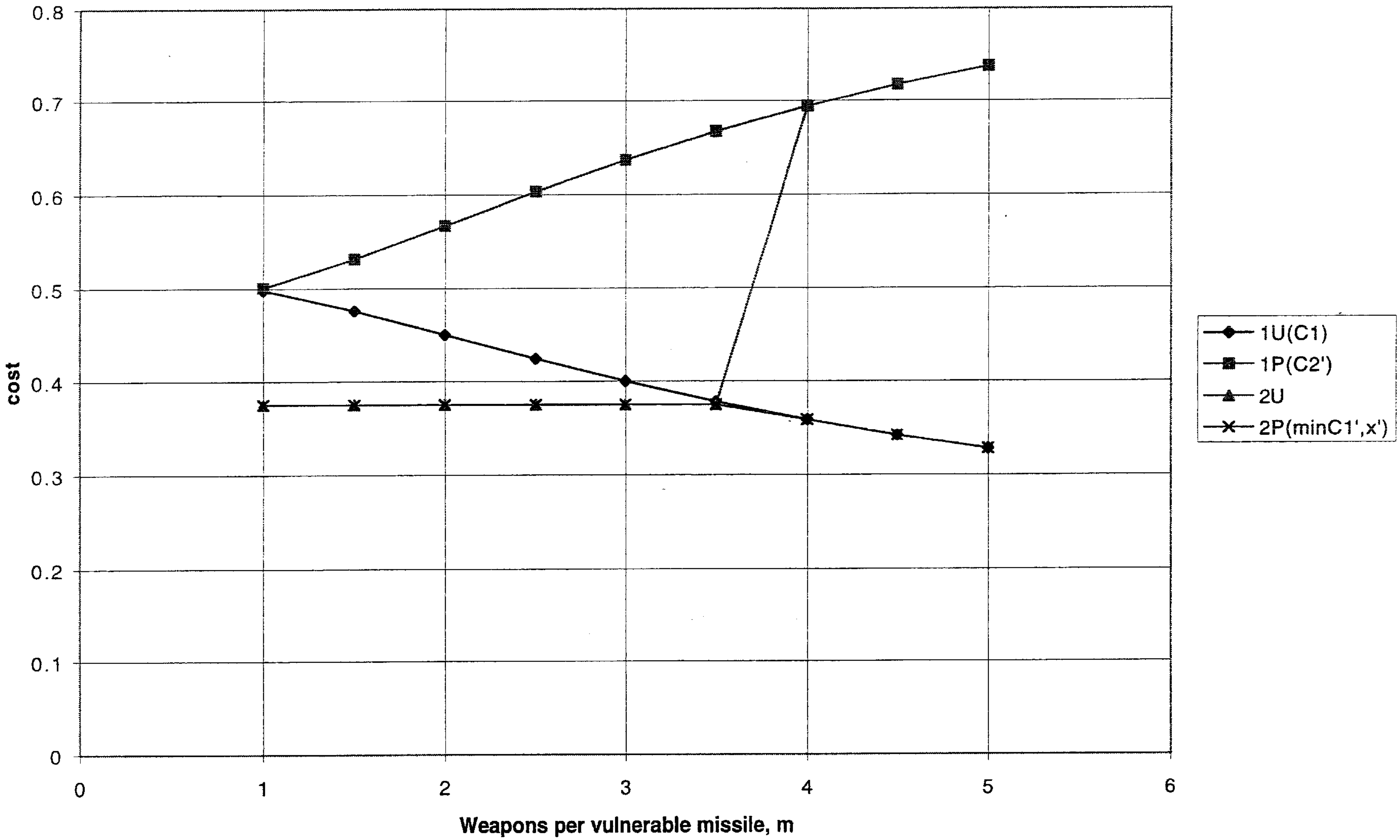


Fig. 6. Cost of node 5 vs weapons per missile

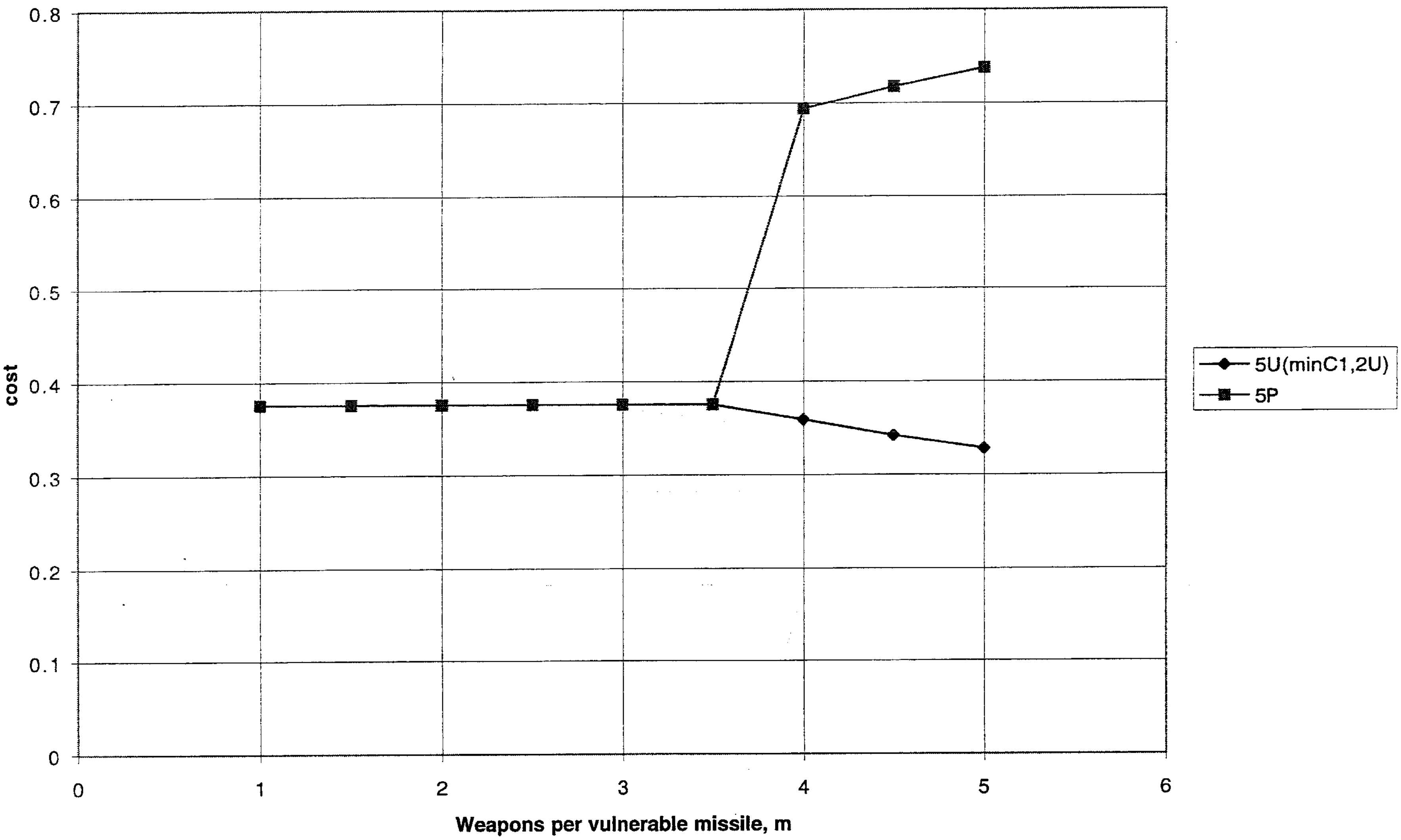


Fig. 7. Cost to P of nodes 5, 6, & 7; (u =0.5)

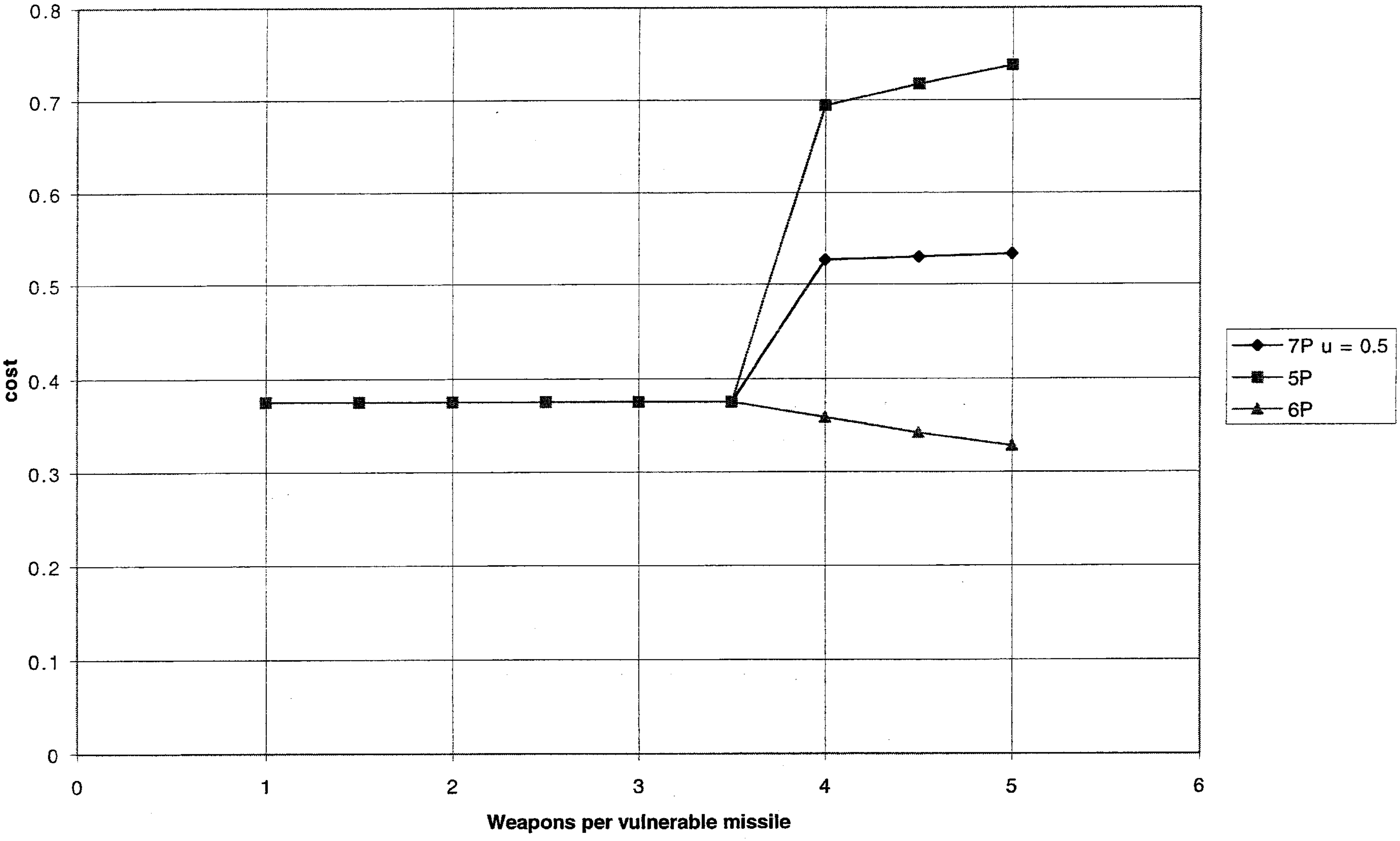


Fig. 8. Optimal allocation f vs period

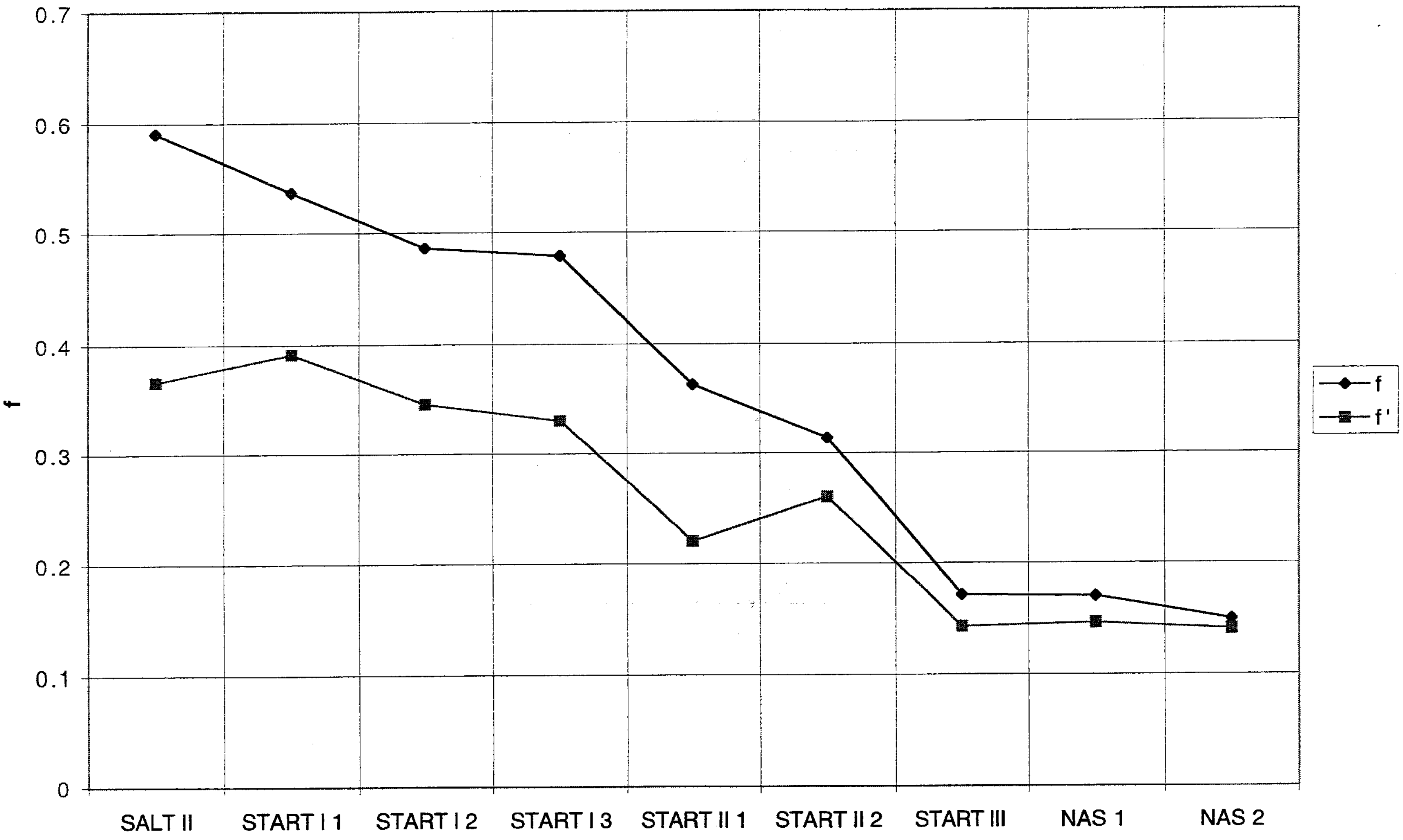


Fig. 9. First and second strike vs period

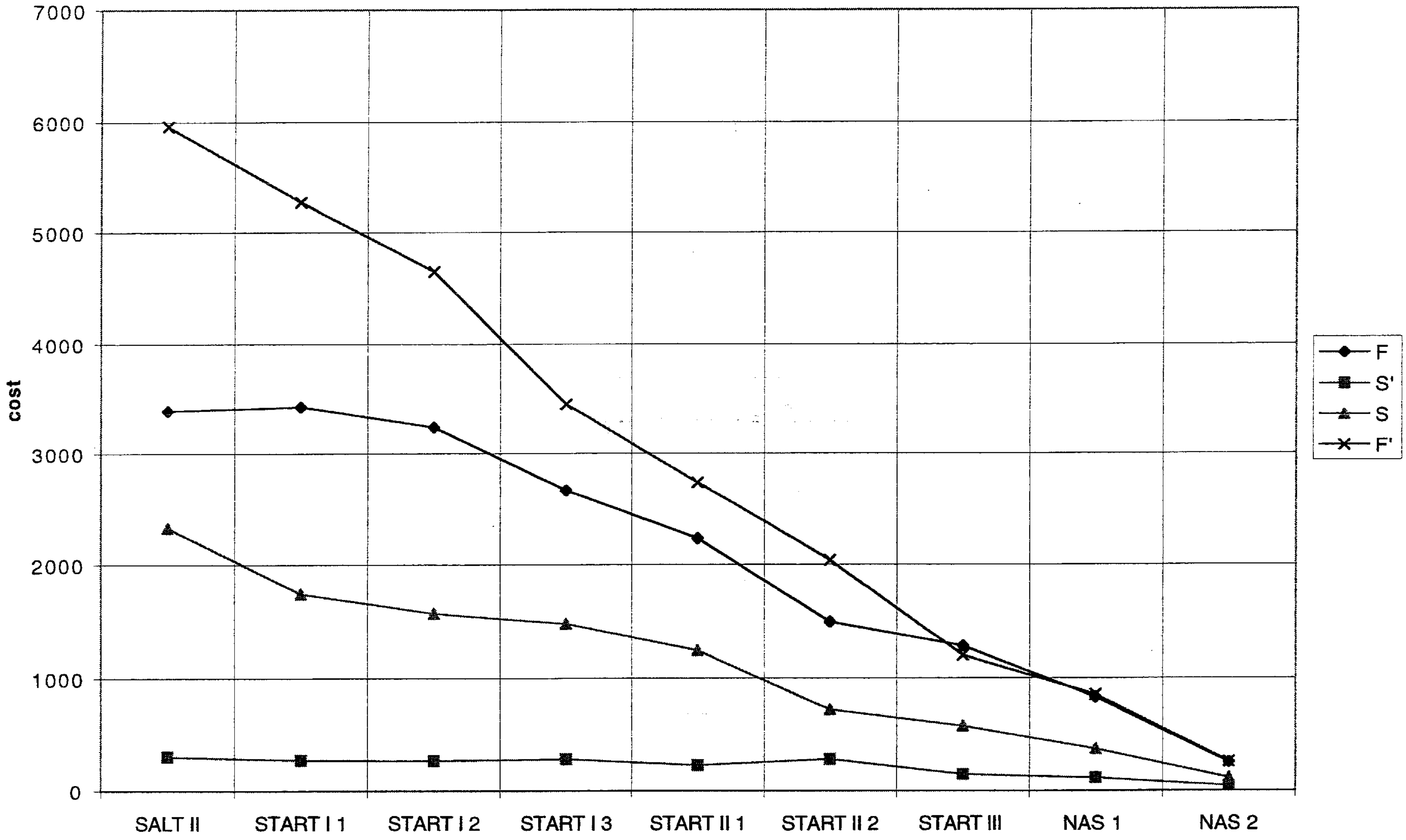


Fig. 10. First and second strike costs C1 & C2

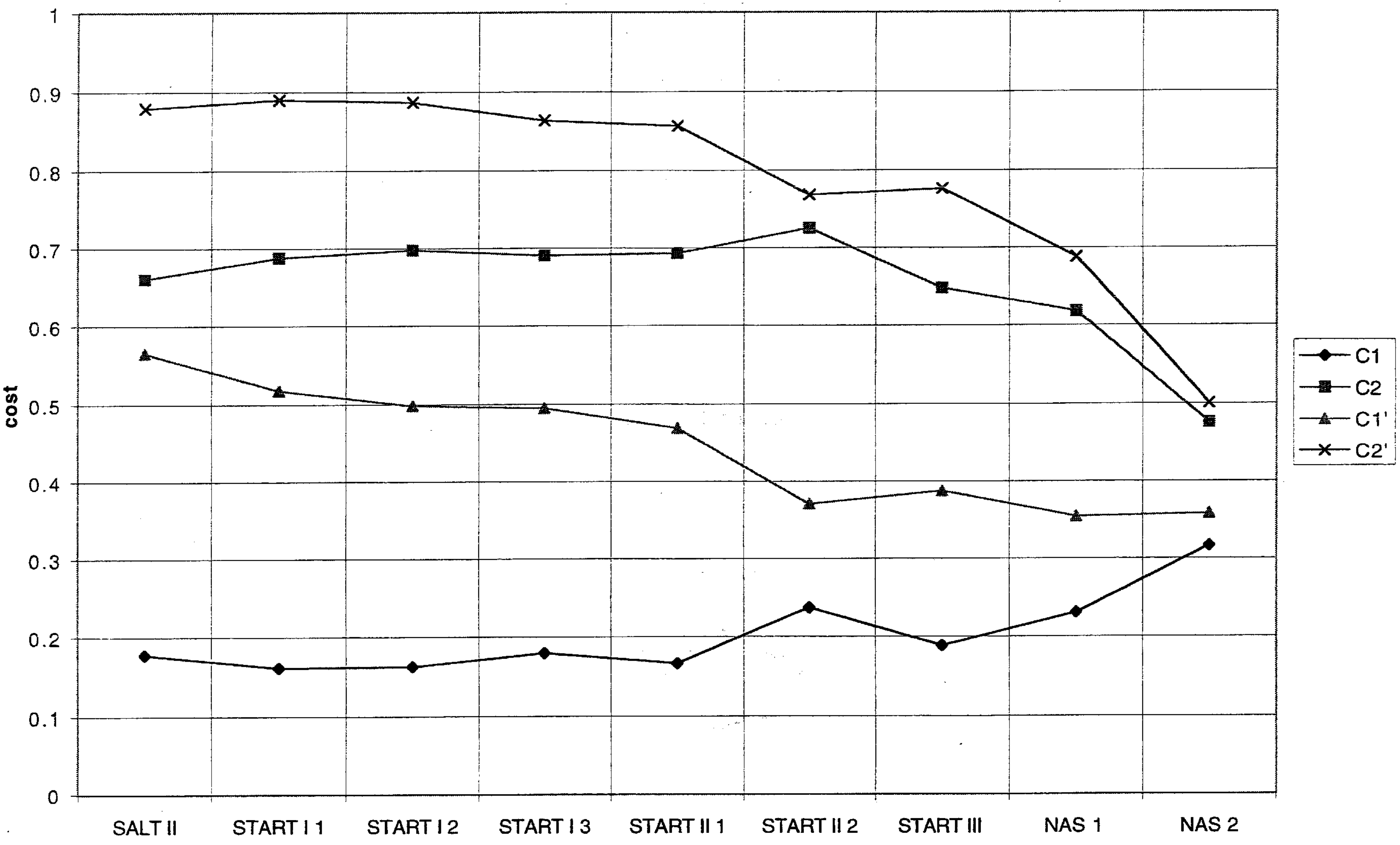


Fig. 11. Costs of Nodes 1 & 2

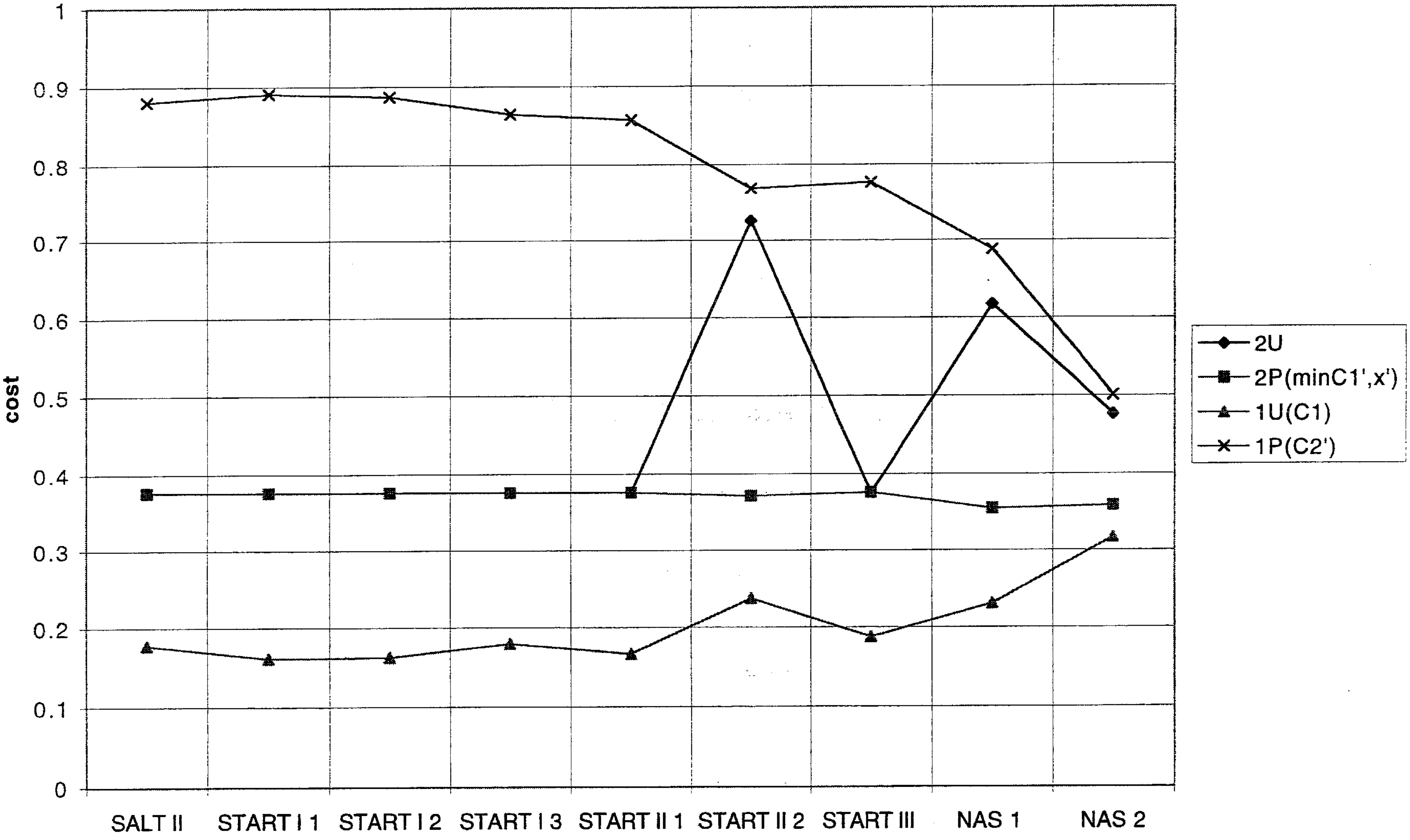


Fig. 12. Cost of node 5

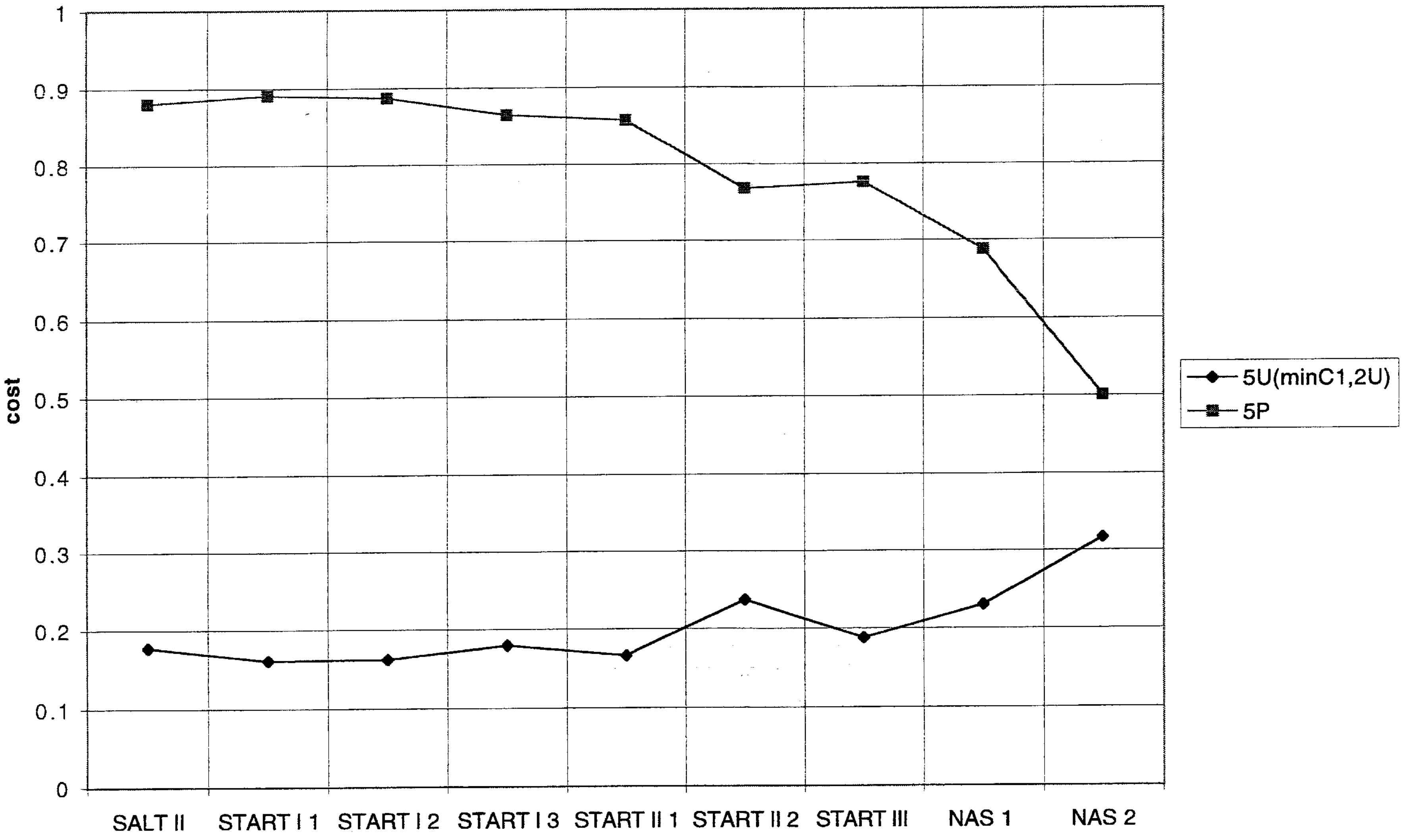


Fig. 13. Cost of nodes 3 & 4

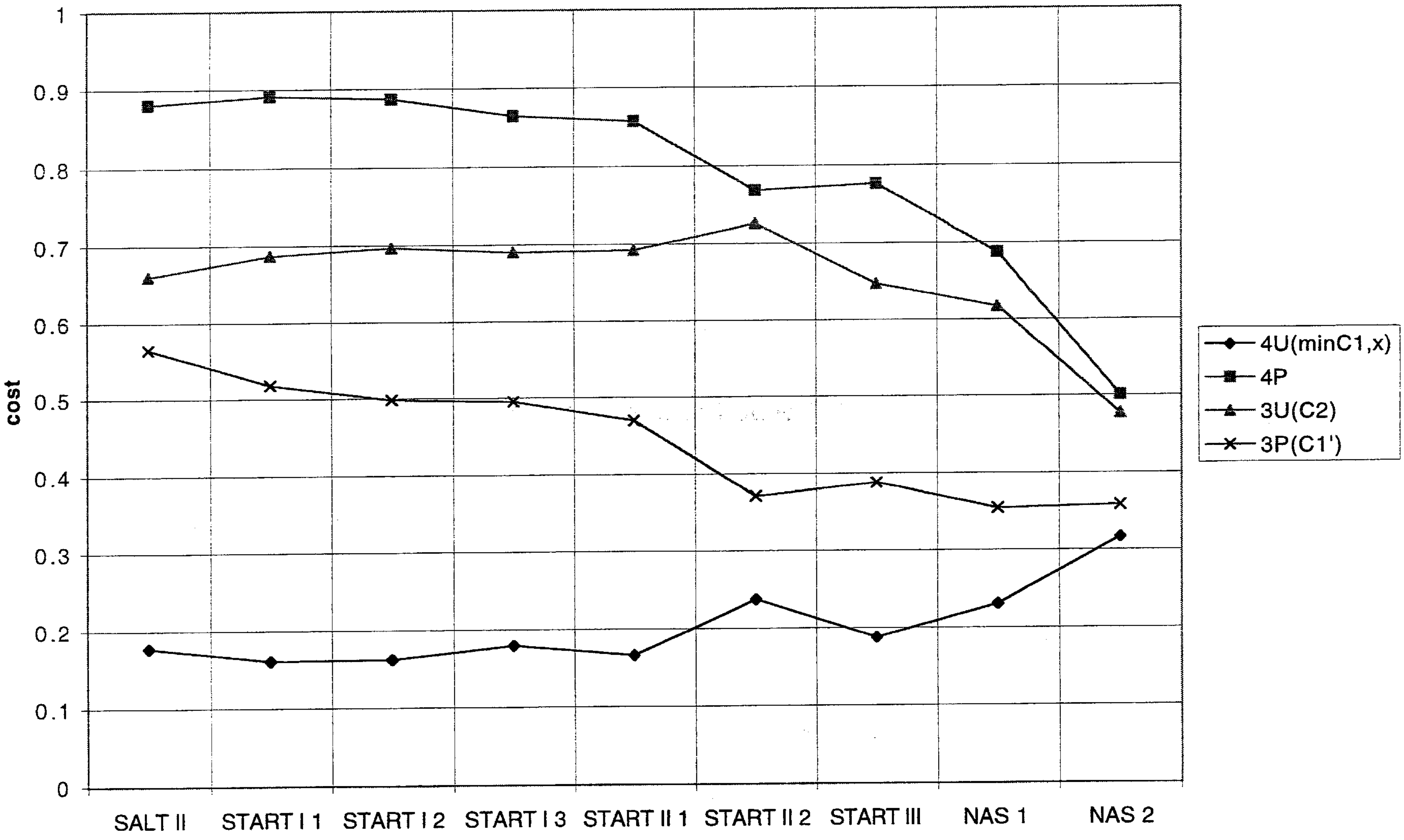


Fig. 14. Cost of nodes 5 & 6

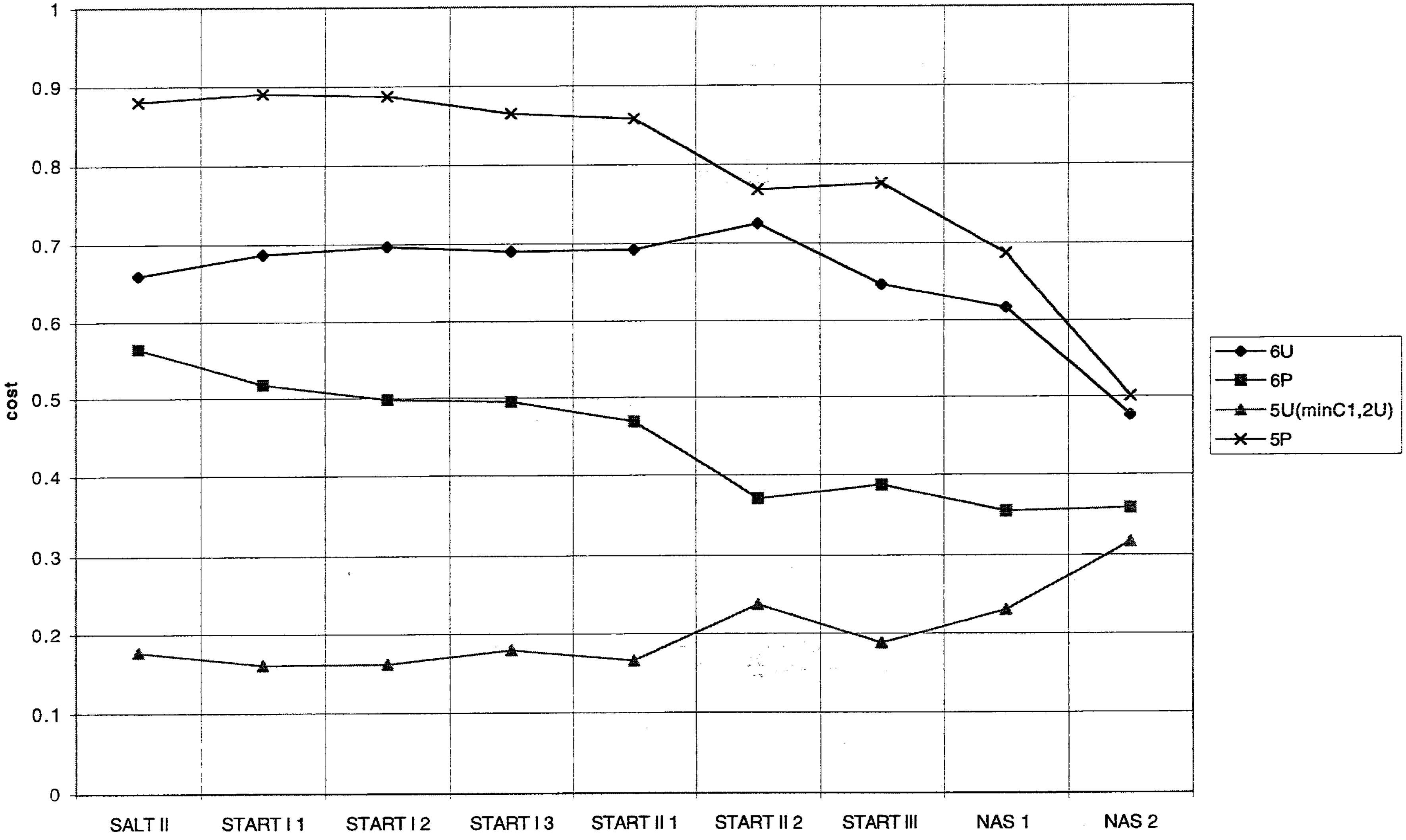


Fig. 15. Cost of node 7 (u=0.5)

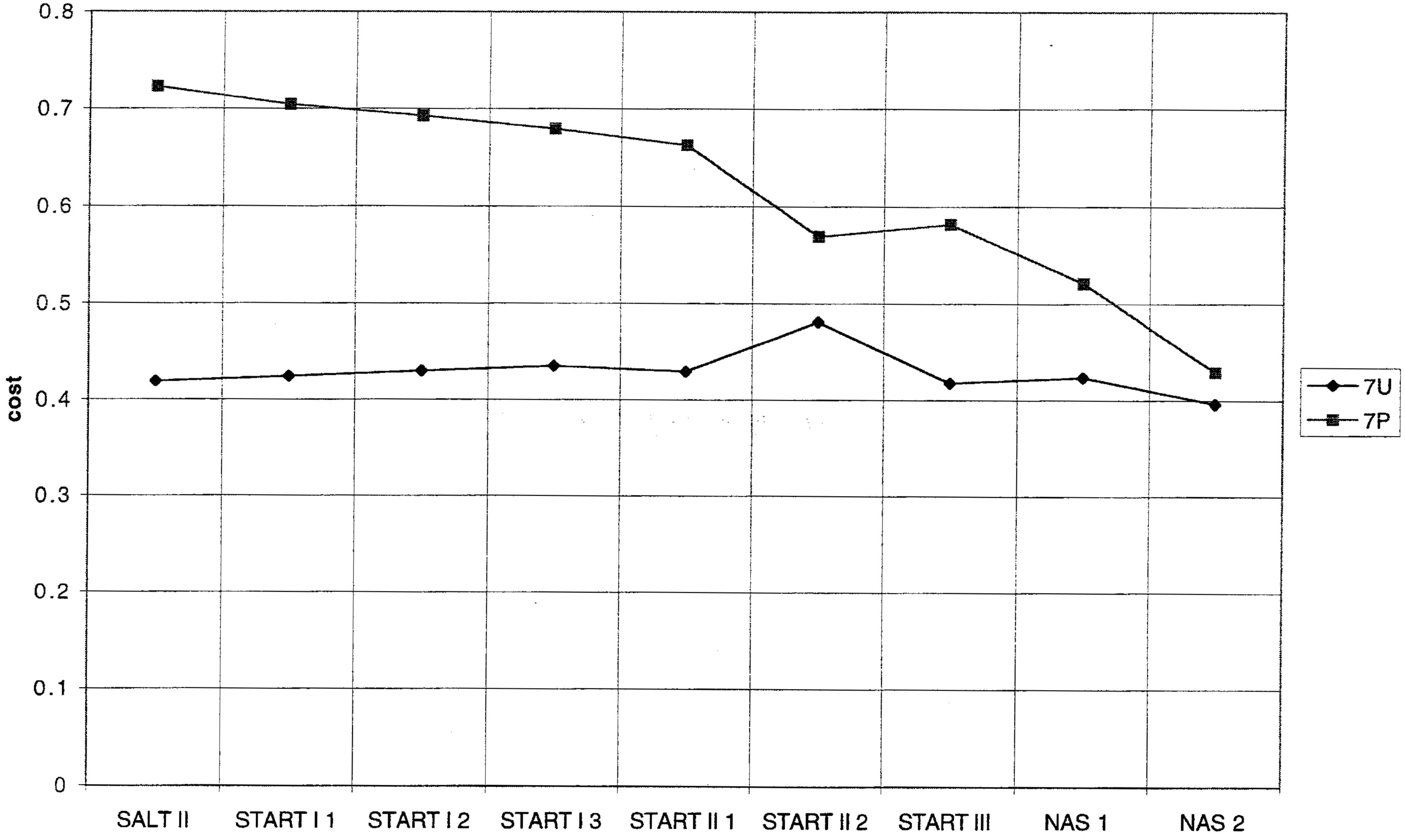


Fig. 16. Breakeven probability by period

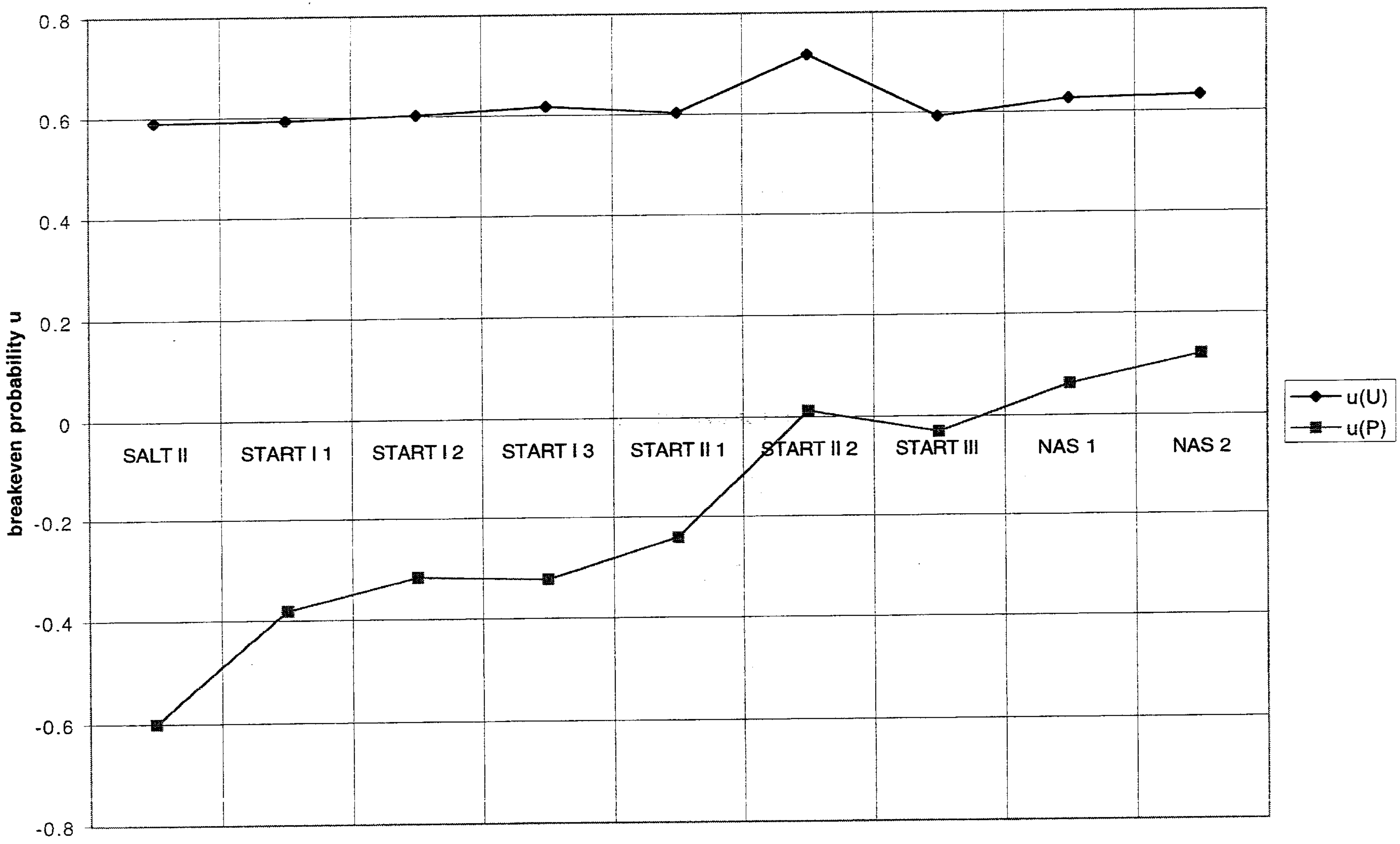


Fig. 17. Weapons Allocations in Trilateral Exchanges

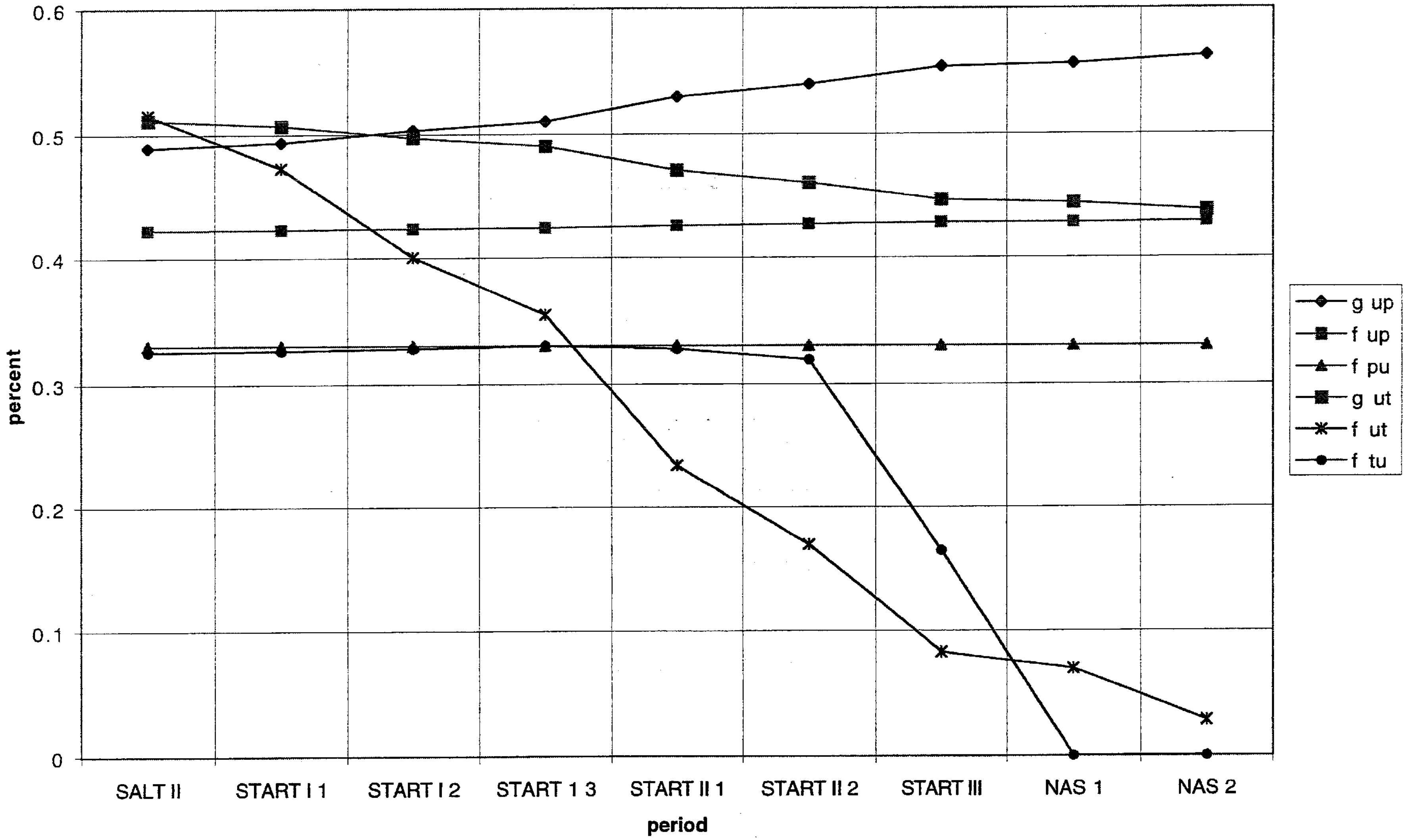


Fig. 18. First strikes

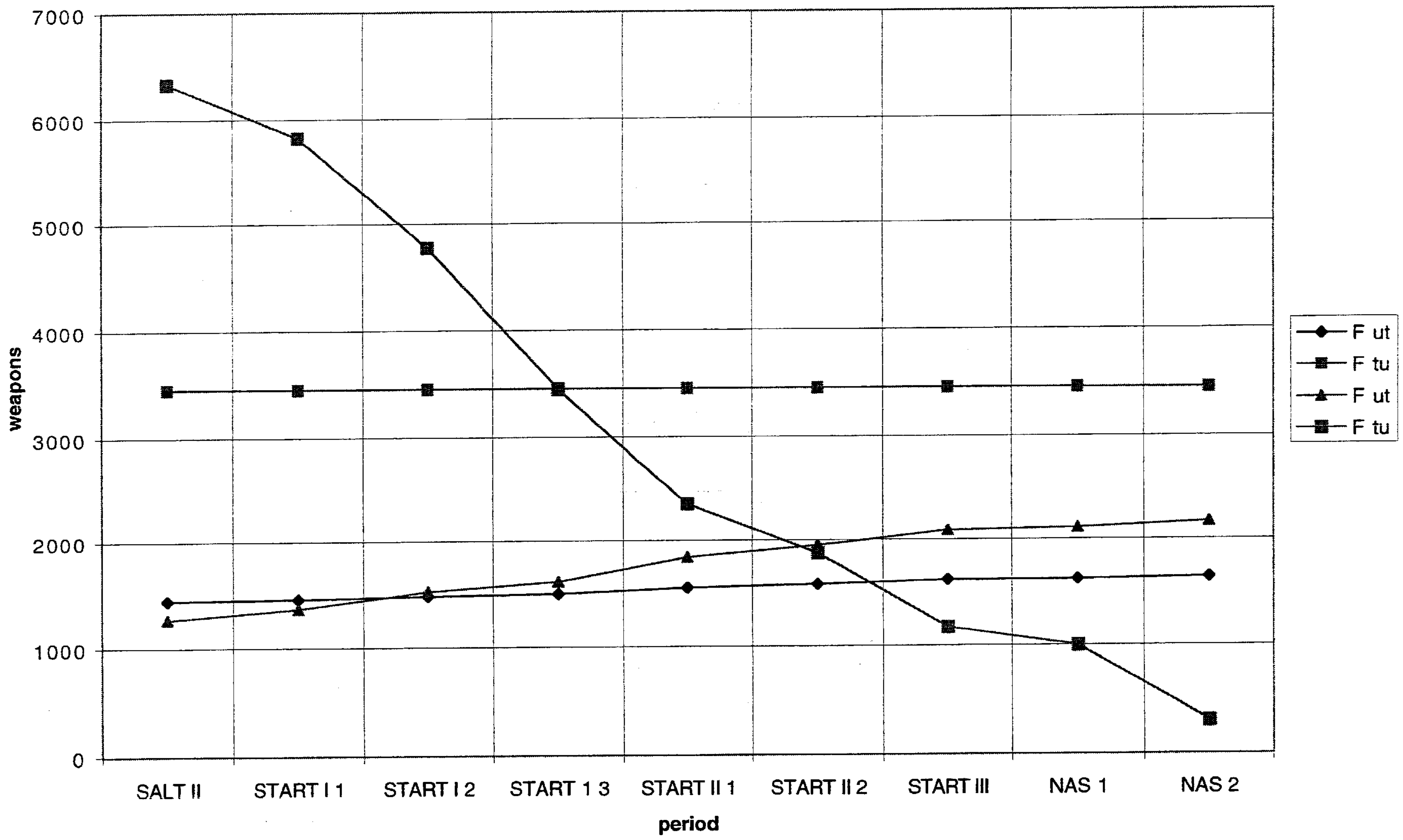


Fig. 19. Second strikes

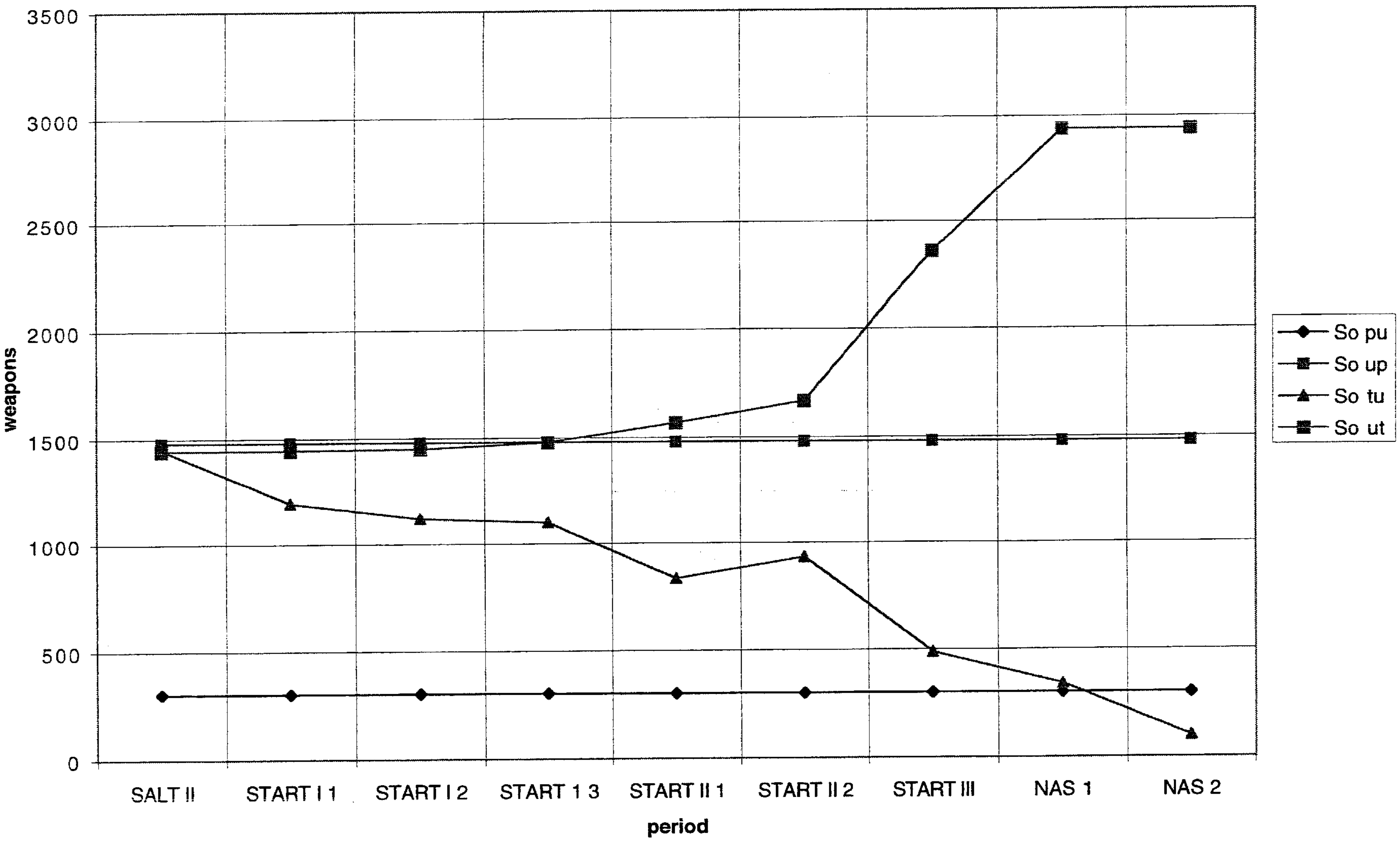


Fig. 20. First strike cost C1

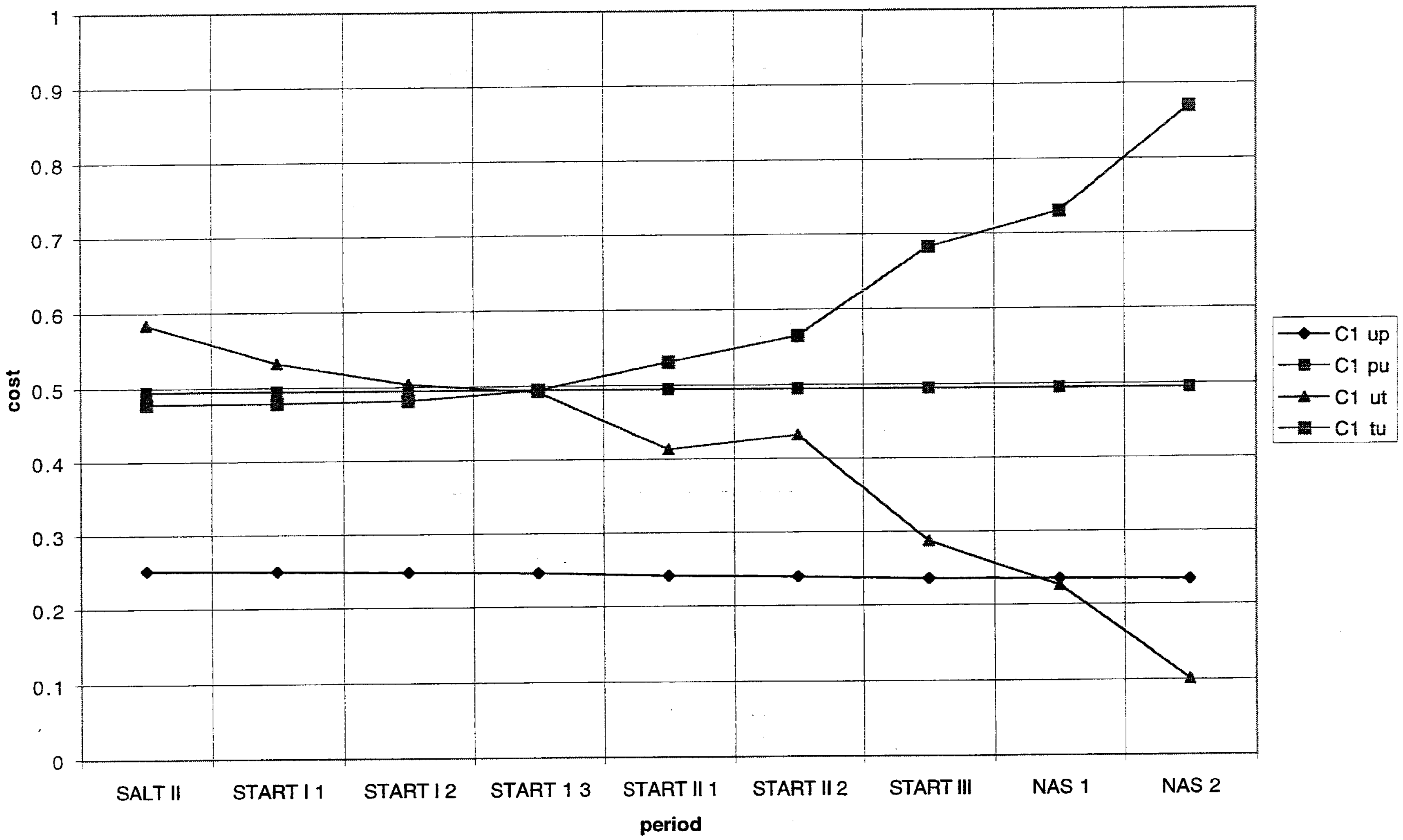


Fig. 21 Cost of node 7 for U &n P

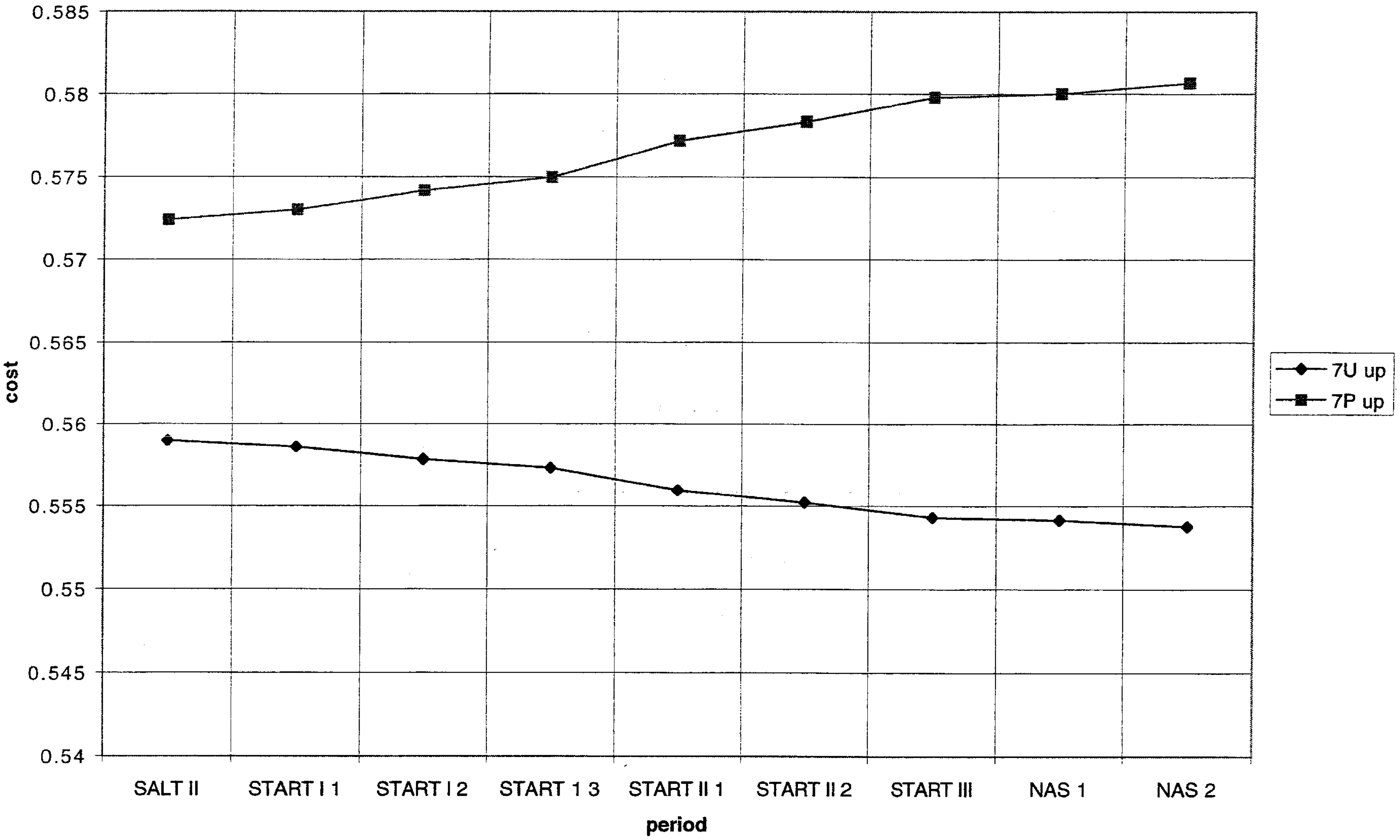


Fig. 22. Cost of nodes 1 & 2 to U & T

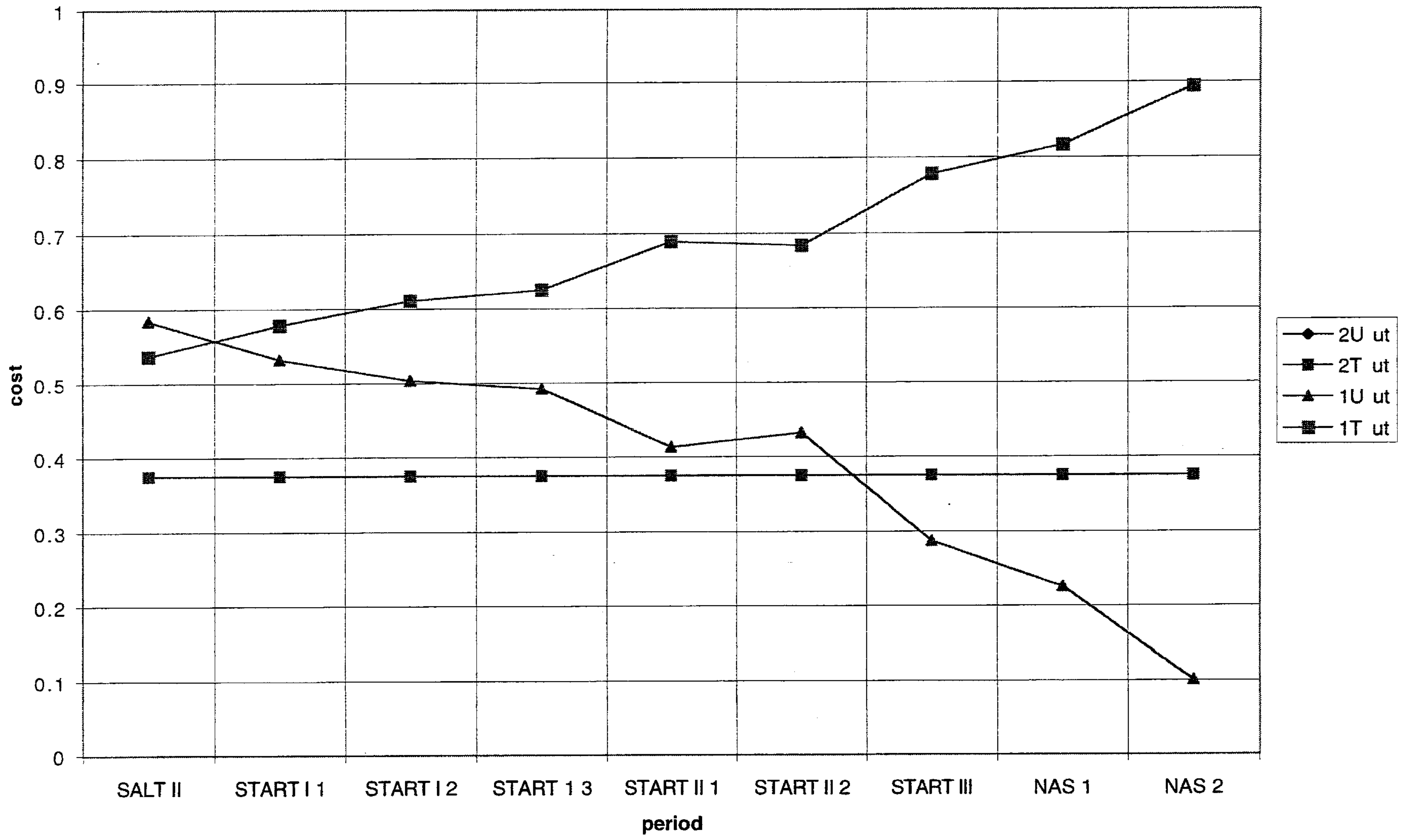


Fig. 23. Cost of node 5 for U & T

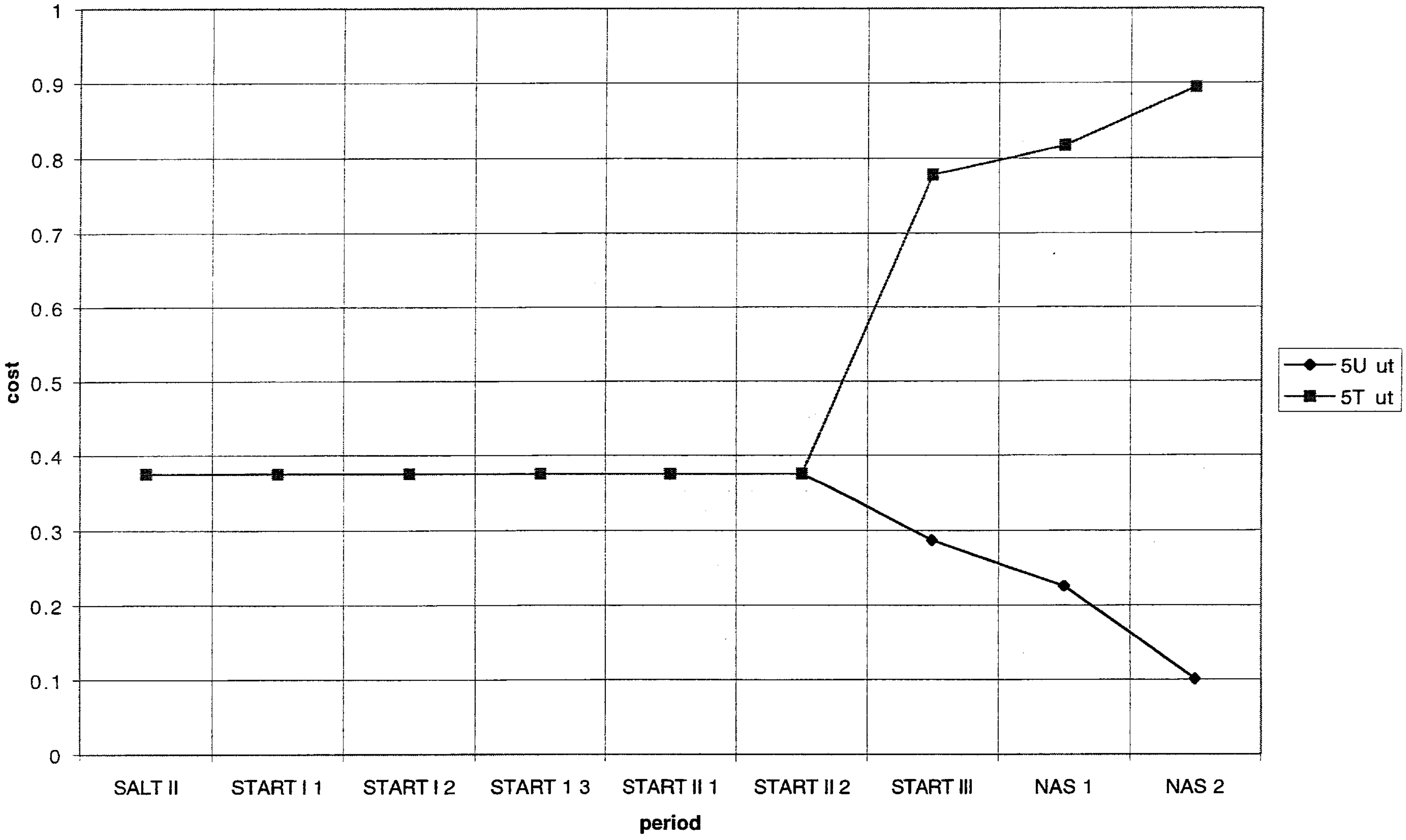


Fig. 24. Cost of nodes 3 & 4 for U & T

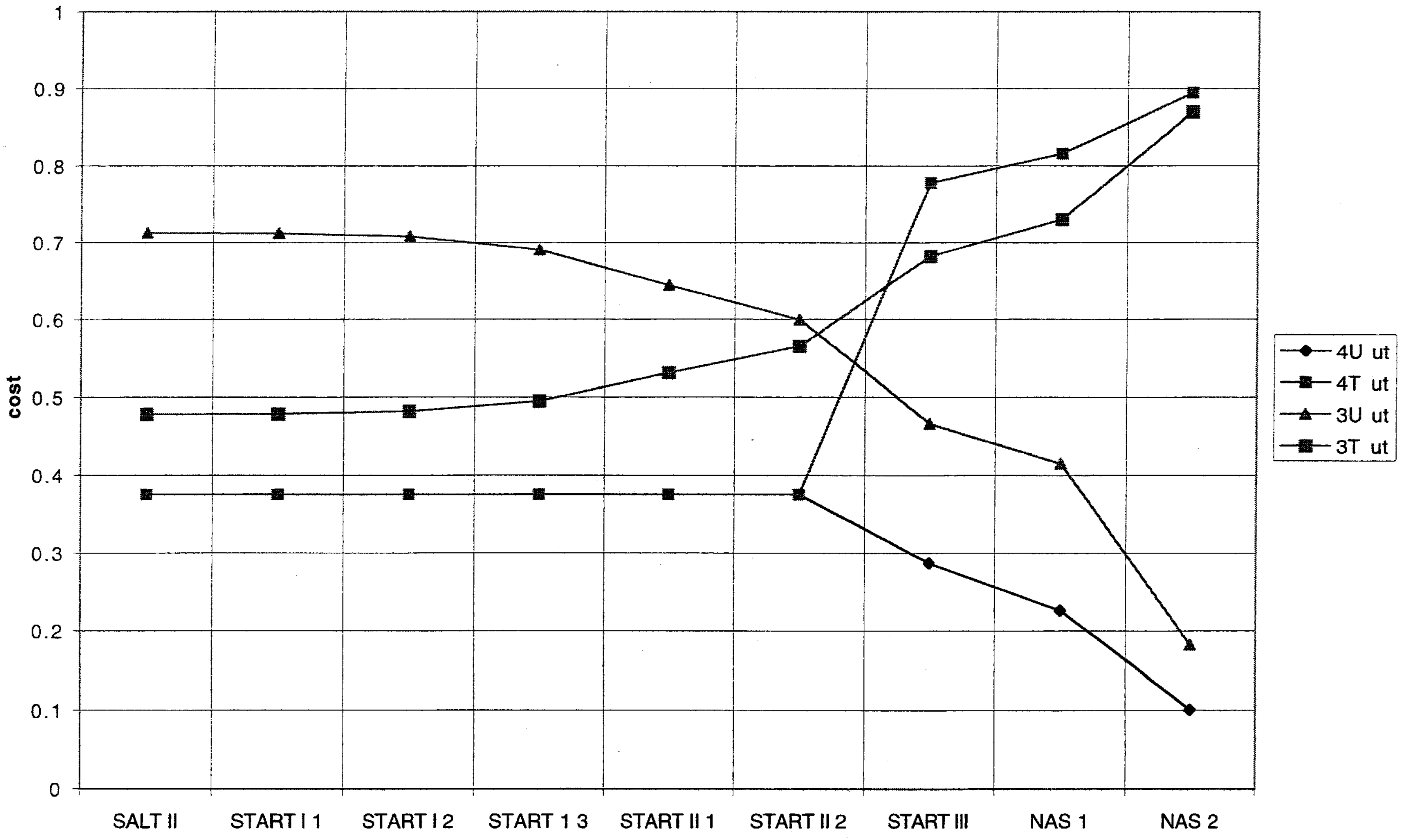


Fig. 25. Cost of node 6 for sides U & T

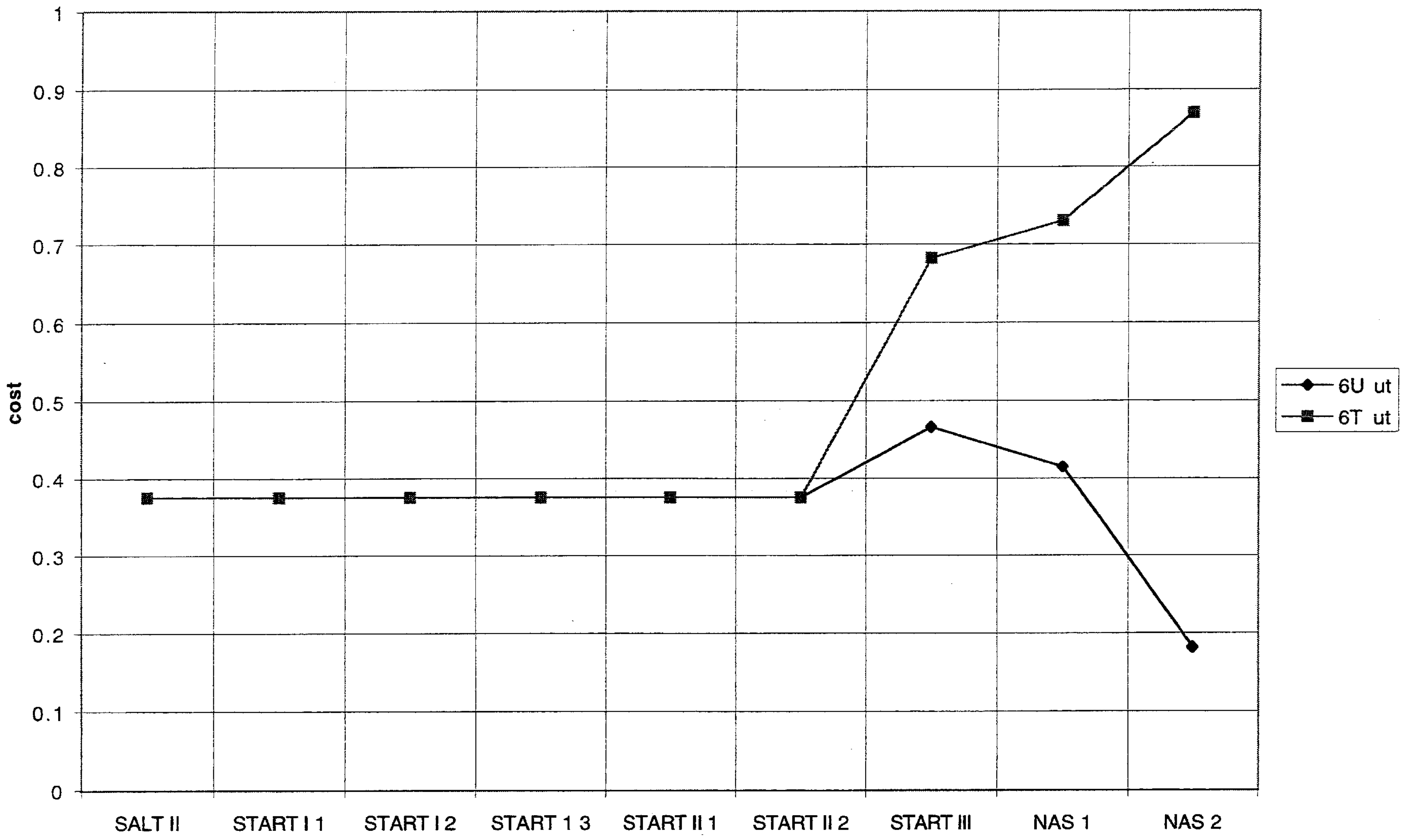
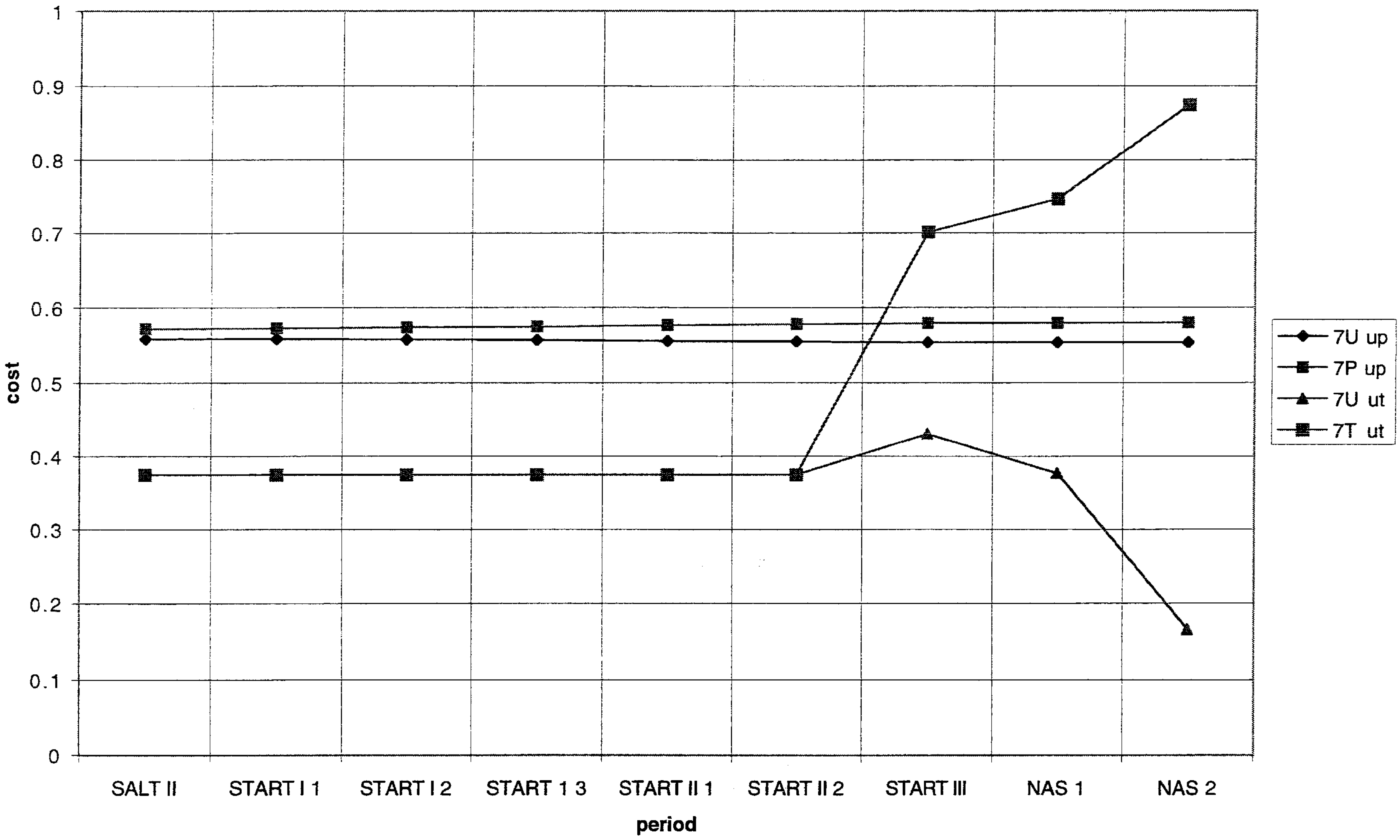


Fig. 26. Cost of node 7 for conflict between sides U & P and sides U & T


cost
SALT II
START I 1
START I 2
START 1 3
START II 1
START II 2
START III
NAS 1
NAS 2
period
7U up
7P up
7U ut
7T ut