Counterterror Intelligence Operations and Terror Attacks

Jonathan S. Feinstein
Yale School of Management

Edward H. Kaplan
Yale School of Management
Yale School of Engineering
Yale School of Public Health

Abstract

We present a formal model of an intelligence agency that must divide its resources between the collection and analysis of information pertaining to terror plots. The model highlights the negative consequences of queues which form when collection exceeds analytic capacity. We incorporate the response of a terrorist organization to the operating characteristics of the intelligence system it faces, and solve for equilibrium strategies for the intelligence system and terrorist organization. Our results demonstrate the importance of properly balancing resources between collection and analysis, and stand in contrast to the observed state of overcollection in US intelligence agencies.

Keywords Intelligence operations · Intelligence collection and analysis · Counterterrorism · Queueing models · Game theory
JEL Classification H56, D29, D73

1 Introduction

Intelligence activities center around two activities, the collection of information and the analysis of what has been collected to identify and interdict potential hostile events (see for example Department of Homeland Security 2011). In this paper, we present a formal model of an intelligence agency that must divide its resources between the collection and analysis of information pertaining to terror plots. We incorporate the response of a terrorist organization, in terms of level of activity and scale of planned attacks, to the operating characteristics of the intelligence system it faces, and solve for equilibrium strategies for the intelligence system and terrorist organization. Our results demonstrate the importance of properly balancing resources between collection and analysis and provide a framework for determining the appropriate allocation balance for a given environment.

The specific issue that sparked our analysis in this paper is the problem of overcollection. In this day there are many channels available for collecting information about potential terror plots, ranging from citizen tips to informants to

a wealth of satellite and internet/computer data sources. Intelligence agencies run the risk of becoming so inundated with information about possible leads that they are unable to keep up and investigate leads in a timely fashion, to determine which are serious threats that need to be investigated in depth and stopped (Blair and Leiter 2010; Priest and Arkin 2010).

When collection overwhelms analysis a queue will form containing initial leads or tips about potential plots that have been collected – but not yet analyzed. Plots in queues represent a real risk – a terror plot in the queue can "blow up" as a terror attack before being analyzed and interdicted. Further, when terrorist organizations gain a sense that an intelligence system has a large backlog of not-yet-analyzed cases, and is being inundated by information, much of it irrelevant to real attacks, they may become emboldened, sensing the system is vulnerable. Thus properly balancing collection activities with analysis is vital for a well-functioning intelligence system.

To address this we issue we build a model of an intelligence system. In our model the intelligence agency has a fixed budget and allocates a certain number of agents to collection activities, for example scanning satellite data or internet traffic data, and the remainder to analysis. The model has several important features that we believe fit with the real world of counterterrorism. First, there are both real plots and "fake"plots or false leads (Department of Justice 2008; Kaplan 2010, 2011; Steele 1989). Real plots take time but, if not interdicted, eventually "blow up" as real terror events, while fake plots never blow up but are discovered to be fake through analysis and discarded. While it would be easiest if fake plots were known to be fake from the start and could simply be ignored, in reality there is no way to know when an initial lead about a plot is collected whether the plot is real or fake. We show that a plethora of false plots has a major impact on how the intelligence system operates and that in this situation overcollection is a serious concern.

Second, we specifically allow for a situation in which the intelligence system becomes overburdened through gathering preliminary information about very many potential plots - some real, many false leads- and has insufficient resources to analyze all the plots in a timely manner – a very real problem in today's intelligence world. In this situation a queue forms: the queue consists of plots about which initial leads have been gathered, and a plot remains in the queue until an analyst picks it up to analyze it.

Third, our model is explicitly strategic so we can explore how terror organizations respond to a counterterrorism agency and its division of resources between collection and analysis. We identify two important links: a terror organization can throw a system into a queue by raising its rate of attacks; and when a queue forms it will take longer (or be less likely) for the counterterror agency to interdict a plot and thus terror organizations will choose to launch larger scale attacks. Thus intelligence system design and operation matter for how terror groups operate.

Our main result is a no queue result: Over the full range of strategic environments we analyze we find that it is always optimal for the agency to manage its resources so that no queue forms. Practically, this means allocating

enough resources to analysis relative to collection. When the agency moves second, responding to terrorist activity, it should balance resources so that all analysts are always busy but no queue forms. When the agency moves first and terrorists are able to respond to its allocation of resources, it is optimal for the agency to invest in even more analysts than the balance point, so that in equilibrium analysts are not always busy. It does this in order to ensure that the intelligence system does not develop a queue: It is better to slightly overinvest in analysts to ensure a queue does not form – if the intelligence agency tries to achieve exact balance when it moves first we show that the terrorists will push the system into a queue mode.

Our results counter what we believe has been the situation for many real world intelligence systems, including US counterterrorism in the past decade. As part of our analysis we calculate the loss in social welfare that occurs in our model when queues form and show that the loss is, for our parameter values, large. When a queue forms terrorists will launch more plots, and larger plots - since on average it will take the intelligence system longer to interdict a plot. On both counts queues hurt social welfare. On the opposite end of allocation, when there is insufficient collection again real plots tend to escape detection and terror groups can plan more and larger scale attacks. Thus the intelligence system must balance resources between collection and analysis. But there is an asymmetry, as our simulation results make clear, and a queue is particularly costly in terms of the social damage of successful attacks. In our concluding section we discuss some possible explanations for why overcollection occurs in real intelligence systems, leading to queues, despite the fact that they may be quite harmful for social prevention of terrorist attacks.

2 Model

2.1 Real and fake plots

The model occurs in continuous time, which makes certain formulas easier to derive. Real and fake plots arise continuously. We focus on a steady-state world in which these rates are uniform over time. While this assumption is made for simplicity, data collected by Strom et al. (2010) support the assumption that terror plots arise fairly uniformly over time. The model is deterministic, but the intelligence system throughput and social outcomes can be interpreted as expected values in a stochastic environment. Fake plots arise at the rate ϕ, where ϕ is taken as exogenous. Fake plots never blow up but rather if a fake plot is not collected and recorded as a plot by the intelligence system it simply evaporates – a lead never collected that went nowhere. However, when a fake plot is collected by the intelligence system it must be analyzed to be recognized as fake. Real plots arise at the rate α where α is chosen by the terrorist groups. We take α as a flow that can be interpreted as translating into the number

of plots per year. Real plots will ultimately blow up as terror events if not interdicted. Real plots are characterized by their scale s. We assume that the time it takes a real plot to blow up as a terror event is linked to its scale, so that larger scale means on average that it takes longer for a plot to hatch as a terror event. Scale can be thought of as the amount of resources, including planning, required to execute the attack. We assume that scale is also linked to expected damage if the plot hatches as a terror event, which we take as average number of casualties. We define social damage below. We assume there is a maximum feasible scale, $s_{\max}$ and that the risk of a real plot blowing up as a terror event in a given instant dt is $\mu\ dt$ where:

$$\mu = \frac{s_{\max} - s}{s_0} \tag{1}$$

and s_0 is a parameter that governs the relationship between scale and mean time to realization. The time from the inception of a terror plot until it is executed follows an exponential distribution with mean duration $1/\mu$. As s approaches maximal scale $s_{\max}$ the risk of a plot hatching at any instant becomes very small with the implication that the mean time until the plot hatches becomes very large. The terror organizations choose s and thus implicitly μ as detailed below. A real plot that is not analyzed and interdicted will eventually (with probability 1) blow up as a terror event.

In our model fake plots arise exogenously as a byproduct of collection. In particular, terrorists do not intentionally plant fake plots. An important extension of our model is to consider the case in which terrorists do generate both real and fake plots, in order to explore how through the use of fake plots terrorists may disrupt intelligence operations around real plots. While recognizing this as a valuable avenue to explore, we do not pursue it in this paper.

2.2 Collection and analysis

The intelligence agency has a fixed budget B which it divides between its two activities, collection and analysis.[1] The primary resource utilization is human skill and time; it is convenient to think of the agency dividing its personnel into two groups, collectors and analysts. Collection brings plots into the system. Examples of forms of collection include scanning satellite data and internet traffic data, reviewing phone wiretaps and emails, undercover agents or informants, and hotlines. The more resources that are devoted to collection the more likely it is that any given plot, real or fake, will be collected in any given instant of time – thus will be more likely to be collected more quickly. Analysts investigate plots that have been collected to determine if they are real or fake and to interdict real plots. We assume analysts work on one plot at a time - our model can be extended to the case in which analysts may switch between plots,

[1] We note that we take our model to apply only to relatively high priority cases, for which analysis is appropriate, thus view our model as describing the allocation of resources within the high priority domain. Rules and resource decisions for low priority screening activities may be different.

but that adds complication. When an analyst picks up a plot he continues to work on it until he determines whether it is real or fake. If he identifies it as fake he discards it and is ready to begin work on a new plot. If he recognizes it as real we assume it is immediately interdicted, and he is ready to begin work on a new plot.

We use a competing risks framework to model how collection and analysis work. First consider the collection process. Assume c resources are devoted to collection. It is convenient to think of c as, for example, the number of internet sites or satellite images that are being scanned so that c measures how many "nets" the agency has out in the world sifting through data. If each net is imagined as covering a specific block of data - so that as c gets larger more blocks are covered – then all plots, real and fake, in a block are collected, but plots that fall outside the range of blocks being covered by the current level of c are not collected. Thinking of c as percentage coverage, it follows that the rate at which fake plots are collected is $\lambda_F = \phi c$, just a fixed percentage of the generation rate. (We note that c is actually resources, not percentages. To clarify this we could introduce a parameter translating from c into rate of collection - in one interpretation ϕ can be viewed as incorporating this parameter. But in our numerical simulations we set total resources at 100 so that c must lie between 0 and 100. Hence it is natural to think of it as a percentage.) For real plots there are two competing risks: either the plot is collected or it blows up as a terror event before being collected. The rate of blowing up is μ. We employ the parameter δ to translate c into the risk of a real plot being collected, so the rate of collection of real plots is δc. Thus the total rate of blow up plus collection is $\mu + \delta c$. Since real plots arise at rate α and each plot must eventually either blow up or be collected, by competing risks the rate of collection must be $\lambda_R = \frac{\alpha c \delta}{c\delta + \mu}$.[2] Finally it follows that $\lambda_F + \lambda_R$ is the rate at which plots in total are collected into the intelligence system.

Next consider the analysis process. When an analyst is working on a fake plot we assume that the time it takes for him to determine it is fake follows an exponential distribution with mean $1/\psi$. When an analyst is working on a real plot we assume the time it takes for him to determine it is real follows an exponential distribution with mean $1/\rho$. However while an analyst is working on a real plot it can blow up; from above, plots blow up following an exponentially distributed duration with mean $1/\mu$. We assume these two processes are independent of one another and recall also that we assume that once an analyst identifies a plot as real it is immediately interdicted. Relevant rates are given by the reciprocals of these means. In our simulations we assume that ψ is larger than ρ as we believe that many fake plots are fairly quickly recognized to be fake. However this assumption is not essential for our analysis.

Denote the total number of analysts by a. The budget condition is then $c + a = B$. In steady-state a_F analysts are working on fake plots at any moment and a_R analysts are working on real plots (note that any given analyst doesn't

[2] It follows that in steady-state there will be a stock A of real plots out in the world at each moment that have not yet been collected. By flow balance the number entering this stock, α, must balance the number leaving, $(\delta c + \mu)A$, which yields $A = \alpha/(\delta c + \mu)$.

know which kind of plot he is working on). These relative numbers depend both on the analysis process, which differs between real and fake plots as described above, as well as the rate at which real and fake plots enter analysis. Above we give the formulas for λ_F and λ_R, the rates at which fake and real plots are collected. Collected plots enter a queue if all analysts are busy; otherwise they move directly into analysis. As it turns out real and fake plots do not move from the queue (when one exists) into analysis at the same rate, as we discuss below. Hence we introduce two additional parameters: η_F is the rate at which fake plots move from the queue into analysis, and η_R is the rate at which real plots move from the queue into analysis. The relationship of these parameters to λ_F and λ_R depends on whether or not a queue of plots forms waiting to go into analysis - we discuss this next. Given η_F and η_R it is easy to derive the formulas for a_F and a_R as essentially a ratio of inflows and outflows. $a_F = \frac{\eta_F}{\psi}$, the rate at which fake plots enter analysis divided by the rate at which they are disposed of. Likewise $a_R = \frac{\eta_R}{\rho+\mu}$ the rate at which real plots enter analysis divided by the sum of the two rates at which they exit: interdiction and blowing up.

Finally, we can compute the success rate of the intelligence system at interdicting real plots. Recall that real plots form at the rate α. Ultimately, through being collected, analyzed and interdicted they are prevented at the rate $\alpha_U = \rho a_R = \frac{\rho\eta_R}{\rho+\mu}$ where the U subscript is used to indicate that this is the rate of unsuccessful attacks. The rate at which real plots succeed and blow up is then given by $\alpha - \alpha_U$. The total damage inflicted by the terrorists on society is modeled as the product of the successful attack rate and the scale of attacks: $(\alpha - \alpha_U) \times s$.

2.3 Queues

There are three different regimes for the intelligence system. If there are sufficient analysts relative to the rate at which new plots are collected than all plots will be analyzed as they are collected. This is the no queue or Light Traffic (LT) condition. There are two possible LT regimes. First there is the condition of balance in which all analysts are always busy but no queue forms. Abstracting from stochastic fluctuations in arrival of new plots, given all parameters of the system, including α and μ (scale) both of which are chosen by the terror groups, there is a specific number of analysts (relative to the number of collectors or total budget B) such that the system is in balance. The second LT regime is when there are more analysts than required for balance, so analysts are not always busy. The third regime is the Heavy Traffic (HT) regime in which a queue forms. This occurs when there are insufficient analysts relative to collectors. When a queue forms the intelligence organization must specify a rule for how plots are drawn from the queue. We focus on the Last In First Out or LIFO rule, which is the best rule for the organization to follow (Kaplan and Feinstein 2011).

We now derive the equations governing the intelligence system and queue formation in these three regimes. We group the two LT regimes together as

the equations are similar.

2.3.1 LT regimes

In Light Traffic regimes newly collected plots flow through immediately into analysis. Thus $\eta_F = \lambda_F$ and $\eta_R = \lambda_R$. It therefore follows that $a_F = \frac{\lambda_F}{\psi}$ and $a_R = \frac{\lambda_R}{\rho+\mu}$. For light traffic to hold true we also require that $a_F + a_R \leq a$. When the inequality holds strictly we are in LT regime 1, idle analysts, and when it holds with equality we are in regime 2, balance. Given the budget B and the collection rate c this condition is:

$$a_F + a_R \leq B - c. \tag{2}$$

Note that a_R depends on λ_R and μ, and λ_R depends on both α and μ as well as c. Because α and μ are chosen by the terrorists, whether or not the system is in LT depends on the terrorists' behavior and therefore is jointly determined with the equilibrium of the game between the intelligence agency and the terrorists. If condition (2) fails to hold the intelligence system is in the Heavy Traffic regime.

Terror plots are interdicted at the rate $\alpha_U = \rho a_R$ and the rate at which real plots blow up is $\alpha - \alpha_U$. At the point of balance we can solve directly for a_R. Substituting in expression (2) for a_R and a_F yields: $\frac{\lambda_R}{\rho+\mu} + \frac{\lambda_F}{\psi} = B - c$ which in turn yields a quadratic equation in c:

$$[\phi\delta(\rho+\mu)]c^2 + [\alpha\delta\psi + \phi\mu(\rho+\mu) + 1]c - B = 0$$

The solution (one root) of this expression yields the value of c that achieves balance of collection and analysis It is tempting to conclude that this is the optimal value for c and that is in fact the case in a nonstrategic world. But in the strategic world of terrorism α is not fixed but may be influenced by the counterterrorism agency's choice of c. As a result the optimal value for c can be solved only as part of the equilibrium of the game between terrorists and the counterterrorism agency. We solve for this equilibrium in the next section.

2.3.2 HT regime

In a Heavy Traffic state a queue forms of plots that have been collected but not yet analyzed. For the model of real and fake plots we are utilizing, LIFO minimizes the expected number of terror events (real plots that blow up) for a fixed number of analysts a, and thus is the optimal queue selection rule to employ. The reason is simple. Real plots eventually blow up whereas fake plots, once they are collected, remain in the system essentially inert. Thus the older a plot in the queue is - meaning it has not blown up – the greater the probability that it is a fake plot. Since it is always preferable for analysts to be working on real plots than fake plots – working on a fake plot is a waste of analysis resources since the plot is fake and will never blow up – it follows that the best selection rule is the one that maximizes the probability that a given plot that is

handed to an analyst is real. Since the probability of a plot being fake increases monotonically with age in the queue the optimal selection rule is for analysts to be handed newly collected plots. It is easy to show that this result also holds for the strategic environment of this paper.

To determine the degree to which the intelligence system interdicts real plots we must determine the throughput of real plots into analysis. Since under LIFO a fraction of newly collected plots move directly into analysis it follows that the relative number of real and fake plots going into analysis will be the same as the relative number of real and fake plots collected: $\frac{\eta_R}{\eta_R+\eta_F} = \frac{\lambda_R}{\lambda_R+\lambda_F}$. This yields $\eta_R = (\eta_R + \eta_F)\frac{\lambda_R}{\lambda_R+\lambda_F}$ and likewise $\eta_F = (\eta_R + \eta_F)\frac{\lambda_F}{\lambda_R+\lambda_F}$. We can translate this into the restriction on the number of analysts using the equations $\eta_R = (\rho + \mu)a_R$ and $\eta_F = \psi a_F$. This yields an expression for a_R:

$$((\rho + \mu)a_R = ((\rho + \mu)a_R + \psi a_F) \times \frac{\lambda_R}{\lambda_R + \lambda_F}$$

Substituting the constraint on the total number of analysts: $a_R + a_F = a$ to substitute for a_F yields an expression for a_R:

$$a_R = \frac{\lambda_R/(\rho + \mu)}{\lambda_R/(\rho + \mu) + \lambda_F/\psi} \times a \tag{3}$$

This formula expresses the number of analysts working on real plots as a fraction of the total number of analysts, with the fraction governed by the rate at which real plots are disposed of (either blow up or are interdicted) compared to the total rate at which plots are disposed of, which includes the disposal of fake plots. A similar expression holds for the number of analysts working on fake plots: $a_F = \frac{\lambda_F/\psi}{\lambda_R/(\rho+\mu)+\lambda_F/\psi}a$.

The rate at which the intelligence system interdicts real plots is then $\alpha_U = \rho a_R$ from above and the rate at which real plots blow up is $\alpha - \alpha_U = \alpha - \rho a_R$. The expression for a_R is increasing in the number of analysts a and it can be shown to be decreasing in the rate of collection c (recall that λ_R and λ_F depend on c - substitution yields this result directly). It thus follows that in a nonstrategic world it is best to increase analysts relative to collection until the queue disappears and the system moves to Light Traffic (Kaplan and Feinstein 2011). We show in the next section that for several different versions of the game between terrorists and the counterterror agency this continues to hold true. Nevertheless despite the thrust of these theoretical results we believe that empirically intelligence systems often operate in the Heavy Traffic regime so that equation (3) is highly relevant.

2.4 Terrorist activity and scale of attacks

Though there are many terrorist organizations in the world, to keep the analysis tractable we assume that there is one large organization - or it can be thought of as a coordinated network of organizations – so there is just a single choice

being made about scale and intensity of terror attacks. Thus we assume that all attacks have the same scale s, which in turns means that all attacks share the same parameter μ governing latency until an attack blows up. Likewise we assume a single cost function for attacks that encompasses all attacks being launched by the organization or network of organizations.

Scale maps into the parameter μ according to the formula $\mu = \frac{s_{\max} - s}{s_0}$. Important for our model is the costs terrorists incur to launch terror plots. We assume that costs are increasing in both rate and scale, and use a flexible linear-quadratic specification to capture this dependence:

$$Costs = c_1 \alpha s + c_2 \alpha^2 s + c_3 \alpha s^2 \tag{4}$$

Note that this expression gives costs as a flow or rate since attacks are being launched continuously. Costs depend linearly on the combined term αs which can be thought of as the total damage rate of terror attacks planned – rate times scale. The additional two terms capture convexities in costs which we expect to arise when the terror organization plans more attacks and larger scale attacks (Feinstein and Kaplan 2010). The second term captures a quadratic dependence on the rate of attacks, with this cost linear in the scale; thus for a fixed scale the cost of planning attacks rises faster than linearly as the rate of attacks rises. This will be due to the scarcity of human resources, both for planning attacks as well as executing attacks, other resources, including financial resources, as well as the greater difficulty of identifying suitable targets as the rate increases. The third term captures a quadratic dependence on the scale of attacks, with this cost linear in the rate of attacks. It makes sense that costs will rise more than linearly as scale rises, since the complexity of planning and executing attacks is surely rising, potentially quite rapidly.

The terror organizations' objective is to maximize the net benefits of terror attacks, which is the benefit accrued from successful attacks minus the costs of attacks. We assume that successful attacks yield benefits proportional to their scale. In addition we assume that all attacks not interdicted execute successfully. We have explored the implications of assuming that some attacks fail, presumably for internal logistical reasons, but the qualitative results are very similar to those we report and thus for brevity we do not present the results. The terror organizations' objective function is thus to maximize:

$$(\alpha - \alpha_u)s - c_1 \alpha s - c_2 \alpha^2 s - c_3 \alpha s^2 \tag{5}$$

We note that the optimal choice of α and s depends on the rate of interdiction α_U and thus on counterrorism operations.

3 Equilibrium

In solving the game between the terror organization and the counterterrorism agency we focus on the scenario in which the agency moves first, dividing its resources between collection and analysis, then the terror organization chooses

its optimal intensity and scale of attacks. We believe this is the most realistic scenario. One reason is that the agency will typically have to fix its collection and analysis resources ahead of time, for example at the beginning of each fiscal year. Equally or more important, both collection and analysis require start-up time and fixed investments: for example satellites for collection must be launched, other kinds of automated collection systems also must be set up, and analysts must be trained. We also consider a second scenario, the opposite case in which the terror organization moves first, choosing its intensity α and scale s of attacks, then the counterterrorism agency chooses how to divide its budget between collection and analysis. Finally, we describe an equilibrium of the simultaneous move game.

The analysis of the second scenario is simpler, hence we begin with a discussion of equilibrium for this case. The intelligence agency will always choose to allocate its resources so as to achieve balance in the intelligence system, meaning all analysts are always busy, but there is no queue. It is easy to see the logic that drives this result. The choices of the terrorists are fixed. Given that, the intelligence system will never want to have analysts idle, since that is a waste of resources. Further, it will never want to allow a queue to form, since in that case some collected cases are discarded (under LIFO) and never analyzed, which means that it would be more efficient in terms of resource utilization to reduce collection slightly at the margin and shift those resources into analysis, eliminating the queue. Thus the only possibility is for the agency to allocate resources to achieve balance. Rolling back, the terror organization chooses its optimal intensity and scale of attacks, recognizing that whatever its choice the allocation of resources by the intelligence agency will be such as to maintain balance in the intelligence system.

The first scenario is considerably more subtle. We analyze this scenario by considering separately the terrorists' choices under Light Traffic and under Heavy Traffic. Whenever they have a choice between these two outcomes they choose the one which yields the greatest net benefits for them. In this scenario, a simple argument by contradiction shows that the equilibrium will not involve balance. We outline this argument focusing on the terrorists' choice of attack intensity α, which is the key factor that can push the system into Heavy Traffic. Suppose that in the equilibrium the intelligence system is in balance. Now consider the terrorists' choice. Within the Light Traffic regime they must prefer balance to any other case since it is their equilibrium choice. This means that the derivative of their net benefit function with respect to α must be zero, or if balance is the corner solution, positive (raising α increases plot rate and therefore if this derivative was negative they would wish to lower α moving into strict Light Traffic with idle analysts). Now suppose they raise α a small amount $d\alpha$. This pushes the system into Heavy Traffic since the rate of plot formation goes up a tiny amount. In turn this adds one new term in the derivative of their objective function, which represents a slight increase in the fraction of plots that avoid interdiction (because they are discarded under LIFO) and thus blow up successfully. In turn, given that all the other terms in the derivative of their net benefit function with respect to α are zero (or at a

corner positive), in fact to second order since they are at a maximum, their net benefit must rise a tiny amount. This shows that their optimal choice will never be one that leads to balance. Holding the agency's allocation of resources fixed, if we start from a low α in which the terrorists' choice leads to Light Traffic and gradually consider raising α eventually, at some value of α, the terrorists will be indifferent between the Light Traffic outcome or the Heavy Traffic outcome – which occurs at with a jump up in α. But it is important to note that their net benefit function will be continuous at this point (albeit with a kink, and a jump in α).

Rolling back, consider the choice by the intelligence agency of how much resources a to allocate to analysis. For any a for which the terrorists would choose an α such that the system would end up in HT, the agency can do better by shifting resources from collection to analysis. We argue this in two steps. First, for a fixed attack rate and scale the agency would always shift resources to analysis since there is overcollection. Second, the terrorists will in turn when resources are shifted lower their attack rate - at the initial attack rate their benefits are now lower since the intelligence system is operating more efficiently, while their costs remain the same, hence they will lower α in order to reduce costs (scale also goes down for the same reason). Conversely, there can never be an equilibrium in which a is so high as to put the system in LT above the point at which the terrorists are indifferent between the LT and HT regimes: the agency can always improve its outcome by decreasing a which increases collection without degrading analysis at all since some analysts were idle. Thus the equilibrium must occur for an agency choice of a^* such that the terrorists end up precisely at the indifference point described above, where they choose an α (and scale s) such that they are indifferent between the LT and HT outcomes. In this case the intelligence agency prefers the LT outcome: for the terrorists the jump into HT increases the rate of attacks and thus social damage but that increase is balanced by an increase in their costs of attacks - but the government cares only about the social damage, not the terrorists' cost, and thus prefers the LT outcome. Thus the equilibrium is for the agency to choose the allocation of resources to analysis such that the terrorists optimal choice leads them to be indifferent between the LT and HT outcomes, and for the terrorists in turn to choose attack rate and scale such that the system ends up in LT. We note that there are issues of stability with this equilibrium and in particular that it is risky for the intelligence agency since if the terrorists choose the attack rate and scale that leads to HT the social outcome is considerably worse. Thus the agency would more likely choose to set the number of analysts slightly above this indifference point.

Lastly, we have identified an equilibrium of the simultaneous move version of our game. This version of our model makes the most sense if each side possesses limited information about the choices made by the other side - it is not realistic to imagine the two sides literally choosing at the same moment, but under some assumptions this might be a reasonable approximation. However, we believe that in reality intelligence systems require more fixed set-up times, especially in today's environment with many automated collection systems, and

that terror organizations are likely to have some information about the intelligence system that they respond to in choosing their attack strategies. Thus this case appears less realistic. A second concern is that equilibria in this game will not have both sides playing pure strategies and thus may be less realistic (although randomized strategies are important in some related contexts; see, for example, Erard and Feinstein 1994). We have identified an equilibrium in which the government plays a pure strategy and the terrorists randomize. We identify this equilibrium piggy-backing off of the equilibrium for the case in which the intelligence agency moves first. In that case the equilibrium occurs in light traffic with the terrorists indifferent between two strategies – one set which leads to the light traffic outcome and a second set, with higher attack rates, that leads to a heavy traffic outcome. For a simultaneous move equilibrium we allow the terrorists to randomize between these two strategies, choosing the rate of attack and scale associated with the heavy traffic outcome with probability θ and the rate of attack and scale associated with the light traffic outcome with probability $1-\theta$. When $\theta = 0$ the intelligence agency would prefer to reduce the number of analysts below a^* to increase collections, while for $\theta = 1$ it would prefer to raise a above a^* in order to move out of heavy traffic. The agency's objective function for each θ is a sum of two terms, what it gets (social welfare) if the terrorists choose to play the heavy traffic strategy, and the other referring to what it gets if the terrorists play the light traffic strategy. Each term is continuous in a and we have just shown that at $\theta = 0$ the agency's optimal choice of a is below a^* while for $\theta = 1$ it's optimal choice is above a^*. It follows that there is some value of θ for which the agency's optimal allocation is exactly a^*. This defines an equilibrium for this game.[3]

For our simulations, for each of the first two scenarios of the game we solve the game by setting up a grid over the possible choices of the first mover and for each choice evaluate the best response of the second mover. For the case in which the agency moves first this is a one dimensional grid since the agency has only a single choice, its allocation of resources between collection and analysis. For the case in which the terror organization moves first the grid is two dimensional and the agency best responses are represented by a surface contour. Once we solve for the best response for each gridpoint we work backwards to the first mover and identify their best strategy. For the simultaneous move game we look for a fixed point, identified above, in which each side chooses to play the strategy we identify given the choice of the other player. In solving the game we evaluate, for each set of moves, whether the agency is working in Light or Heavy Traffic, and if it is operating in Heavy Traffic we impose the LIFO rule for selection of cases from the queue (equivalent to discarding a fraction of new plots).

For the first two scenarios we consistently identify a unique equilibrium in pure strategies, which is the focus of our presentation of results in the next section. We have not explored mixed strategies for these scenarios though as noted above in other kinds of detection games these can be important. For the

[3] For our simulation parameters the equilibrium occurs at $\theta = 0.0157$.

simultaneous move game we believe all equilibria have at least one side playing a mixed strategy; we believe the equilibrium described above is most realistic, with the terrorists playing a mixed strategy and the government playing a pure strategy.

4 Simulation results

Table 1 lists the values for all parameters. Note that for the baseline results the convexity parameters of the cost function are quite small. We choose parameters such that the ratio of the collection rates of fake to real plots is large, on the order of 100 to 1, depending on the exact terrorist strategy; we believe this is realistic, indeed it could be set even higher. We use real and fake analysis parameters such that it takes on average one week to determine that a fake plot is fake and on average 3 months to determine that a real plot is real and interdict it. We normalize agency resources to 100 to be divided between collection and analysis. If this corresponds to human capital, at a cost of $300,000 per individual total per year it corresponds to a budget of $30 million. Table 2 later presents some robustness checks on our results for different sets of parameters. For comparison with our results, consider the optimal choice of scale and intensity of attacks for the terrorist organization when there is no counterterrorism and hence all plots succeed, for the base case parameters in Table 1. In this case where $c_1 = c_2 = c_3 = c$, $\alpha^* = s^*$ by symmetry, and the solution is $\alpha^* = s^* = (1-c)/3c$. For the case of $c = 0.01$ that we use for our main results, $\alpha^* = s^* = 33$. This corresponds to an average of 33 attacks per year with an average casualty rate of 33 per successful attack. These are large values – when we introduce counterterrorism they are reduced very substantially.

[Table 1 near here]

Figures 1, 2 and 3 present results for the strategic scenario in which the government moves first. Figure 1 depicts the optimal terror organization payoffs for the Light Traffic and Heavy Traffic regimes as a function of the counterterrorism agency's allocation of resources between collection and analysis, specifically the value of a. The point where the two lines cross, which is approximately at $a = 37.685$, is where the terror organization is indifferent between the two regimes in terms of payoff, which is in fact the equilibrium for this scenario as we discussed in the previous section. Figure 2 shows the terrorists' optimal attack rate α and scale s choices for the LT and HT regimes, again as a function of the agency's allocation of resources. Note that at $a = 37.685$ which is where the terrorist's shift from their LT to their HT choices, there is a discontinuity. As a increases through this point the terrorists' optimal attack rate falls discontinuously, and their optimal scale rises discontinuously. The equilibrium values at this LT point are $\alpha = 7.693$ and $s = 6.434$. In the context of our model this corresponds to launching between 7 and 8 attacks on average per year, with

average casualties per successful attack slightly over 6. Thus in our model a public expenditure of $30 million enables the interdiction of approximately 22 attacks per year and a savings of perhaps more than 100 individuals. However, we stress that we do not have good cost figures for terror organizations and if costs are higher the actual rate of attacks will be lower.

[Figure 1 near here]
[Figure 2 near here]

Figure 3 depicts the terror organization's optimal payoff as a function of the agency's allocation of resources, over the entire range of possible allocations, from $a = 0$ to $a = 100$. It is noteworthy that the terrorists' payoff diminishes quite steeply to the left of the equilibrium as the allocation of resources to analysis increases, and is quite flat to the right of this point, finally rising steeply when most resources are being allocated to analysis on the far right of the figure. Figure 4 depicts the social damage function facing the intelligence agency, again as a function of its allocation of resources. The curve in Figure 4 is similar to that in Figure 3, reflecting the fact that when the terrorists achieve a high payoff there is high social damage. However, this is not a zero sum game and the curves are not equivalent, because the terrorists also take into account the costs of attacks to their organization whereas the intelligence agency, having a fixed budget, is concerned only with minimizing the social damage of successful attacks. We note that at the agency's optimal choice the system is in LT in the equilibrium (recall the terrorists are indifferent between LT and HT and thus willing to play this equilibrium) and there is a slight excess number of analysts over balance – analysts are idle slightly less than 1% of the time. At the equilibrium the government sets $a = 37.65$ as its optimal allocation of resources to analysis, and the interdiction rate is approximately 77%. This is consistent with the data in Strom *et al.* (2010) for interdiction of terror plots in the United States. Social damage at the optimum is 11.295, which refers to the expected casualty rate from successful attacks.

[Figure 3 near here]
[Figure 4 near here]

The shape of the curves in Figures 3 and 4 demonstrate how important it is for the intelligence agency to maintain something close to balance in its allocation of resources between collection and analysis. When insufficient resources are allocated to analysis a queue forms, and this leads to high social damage. Why is a queue so harmful in terms of social damage? When a queue forms two effects work against the counterterror agency to increase social damage: (i) a smaller fraction of real plots make it to analysis (recall that under LIFO a percentage of collected plots are simply discarded); and (ii) as the rate of collection increases the mix of plots being collected becomes increasingly weighted towards false plots – collection of false plots increases linearly with collection at the rate ϕ whereas collection of real plots is given by $\frac{\alpha c \delta}{c\delta+\mu}$ which increases less than proportionately as the rate of collection c increases. These two

effects together reduce the effectiveness of the agency's interdiction efforts, thus increasing terror success. Note also that as the number of analysts falls from the equilibrium point and a queue forms attack scale is much larger – when there is a queue it takes longer on average for the agency to interdict a real plot and thus it pays the terror organization to plan larger scale attacks, which take longer on average to execute but generate more damage. Conversely, if the agency overallocates resources to analysis it is simply wasting resources, since analysts are idle. While in the equilibrium it must tolerate having some analysts idle, having more than this minimum necessary number idle is wasteful – collections are simply too low with the result that a higher percentage of plots succeed. The far right and far left points in these figures correspond to the case of no effective counterterrorism, in which case the terror organization sets $\alpha^* = s^* = 33$, values that are far above the equilibrium values for these variables.

In the opposite case in which the counterterrorism agency moves second, the agency *always* chooses the number of analysts so as to achieve balance – all analysts are always busy but no queue forms. The terror organization chooses an attack rate of just over 7.5 and a scale of attack of approximately 6.5. The government sets $a = 37.5$. The interdiction rate is approximately 77%, the same as for the strategic scenario in which the agency moves first. Social damages are 11.11, which is slightly below the level for the first scenario. Thus it is slightly better for the counterterror agency to be able to move second. The reason is that in this case it can ensure balance, which is its most efficient use of resources and attack intensity is slightly lower.

5 Conclusion

Our model and analysis leads to a clear policy prescription: Invest in sufficient analysts to avoid queue formation. It is striking that this conclusion stands in direct contrast to the observations of most experts on actual practice in the field of counterterrorism. As recounted in the introduction, queues seem to be commonplace in the intelligence world, especially post 9/11.

Why is overcollection and hence queue formation occurring? While a full exploration of this issue is beyond the scope of this paper, we offer a few possible explanations as some concluding thoughts. In the United States many intelligence agencies are organized around collection activities. Examples include the National Geospatial Agency, engaged in aerial surveillance, the National Reconnaissance Office, responsible for satellite-based collections, and, less transparently but none the less relevant, the Central Intelligence Agency, a centerpoint for human-based collection. For an agency whose mission centers on collection, it is certainly plausible that in some cases collection becomes an end in itself, so that productivity and results are measured in terms of collections. In such a situation there may be a natural bias towards overcollection in terms of the productivity of the intelligence system as a whole in using what is collected - relative to the re-allocation of some resources to analysis. In principle the Of-

fice of the Director of National Intelligence (ODNI) is charged with balancing collection with analysis, in terms of the intelligence system as a whole. Indeed the 9/11 Commission Report makes essentially this very statement. In its call for establishing the ODNI it states: "The National Intelligence Director would submit a unified budget for national intelligence that reflects priorities chosen by the National Security Council, an appropriate balance among the varieties of technical and human intelligence collection, and analysis" (The National Commission on Terrorist Attacks Upon the United States 2004:412). The drive towards overcollection based in the mission of individual agencies can be thought of as an agency problem, in which individual agencies seek to maximize their budgets and performance evaluations by Congress and others, rather than making decisions from the viewpoint of maximizing their contribution to safeguarding national security. Because their collection activities are more readily documented and tied to specific programs for which they have primary responsibility they focus on these and may well overcommit resources to them. Based on the results we have presented, which are admittedly based on a stylized model, as well as the anecdotal evidence, an important further step is to model these agency relationships and evaluation schema and consider policies and evaluation approaches that could promote a better balancing of collection with analysis.

Several additional factors contribute to overcollection. Overcollection seems to have been exacerbated by the advent of collection technologies, notably satellite and digital data sources. The amount of data collected is simply overwhelming relative to the kinds of data available 50 years ago and longer. While these data are obviously a huge boon for intelligence, it also can readily swamp analytic capabilities, which are still far more human-resource intensive. There are two further factors related to this one. The immense growth in intelligence data collection was most likely not fully foreseen in the wake of 9/11. Thus the problem of overcollection is in part an unintended consequence of well-intentioned and for the most part well-grounded decisions about intelligence resource allocation. Finally, there may be some psychological bias at work, in which humans ultimately responsible for security prefer to believe that they have collected as much information about potential terror plots as is possible, even while recognizing that some of that data may never be properly analyzed and therefore may not be of much help in preventing terror attacks. People may feel better imagining they had the data but didn't quite get around to analyzing it, rather than feeling more exposed because the system had never even collected the relevant data.

Moving forward requires recognizing the importance of balancing collection with analysis, building richer, more exact models to evaluate the trade-offs and relationships between the two, and meshing theory with empirical analysis.

Acknowledgment

The authors gratefully acknowledge the Daniel Rose Fund supporting the Technion-Yale Initiative in Homeland Security and Counterterror Operations

Research. We thank an anonymous referee and participants in the 2011 UT Dallas Terrorism and Policy Conference for helpful comments.

References

Blair, D.C., & Leiter, M.E. (2010). *Intelligence reform: the lessons and implications of the Christmas Day attack*. Testimony before the Senate Homeland Security and Governmental Affairs Committee, January 20, 2010, Washington, DC.

Department of Justice (2008). *The Federal Bureau of Investigation's Terrorist Threat and Suspicious Incident Tracking System*. Office of the Inspector General, Audit Report 09-02, United States Department of Justice, Washington, DC.

Department of Homeland Security (2011). *National Terrorism Advisory System: Public Guide*. United States Department of Homeland Security, Washington, DC.

Erard, B., & Feinstein, J.S. (1994). Honesty and evasion in the tax compliance game. *RAND Journal of Economics*, *25*(1), 1-19.

Feinstein, J.S., & Kaplan E.H (2010). Analysis of a strategic terror organization. *Journal of Conflict Resolution*, *54*(2), 281-302.

Kaplan, E.H. (2010). Terror queues. *Operations Research, 58*(4), 773-784.

Kaplan, E.H. (2011) Intelligence Operations Research. *Operations Research* (in press).

Kaplan, E.H., & Feinstein, J.S. (2011). *Intel Queues* (draft paper in review).

Priest D. & Arkin W.M. (2010). National Security Inc. *Washington Post* (July 20) A1.

Steele J.M. (1989). Models for managing secrets. *Management Science, 35*(2), 240-248.

Strom K., Hollywood J., Pope M., Weintraub G., Daye C., & Gemeinhardt D. (2010). *Building on clues: examining successes and failures in detecting US terrorist plots, 1999-2009*. Research Triangle Park, NC: Institute for Homeland Security Solutions.

The National Commission on Terrorist Attacks Upon the United States (2004). *The 9/11 Commission Report*. New York: W.W. Norton & Company, Inc.

Parameter	Symbol	Base Case Value
Annual Real Plot Detection Rate	δ	0.3
Annual Fake Plot Collection Rate	ϕ	30
Annual Interdiction Rate per Real Plot	ρ	4
Annual Discard Rate per Fake Plot	ψ	52
Intelligence Budget	B	100
Scale to Duration Conversion Constant	s_0	100
Maximum Real Plot Scale	$s_{\max}$	100
Terrorist Cost Coefficients	c_1, c_2, c_3	0.01, 0.01, 0.01

Table 1: Parameter values employed in simulation examples.

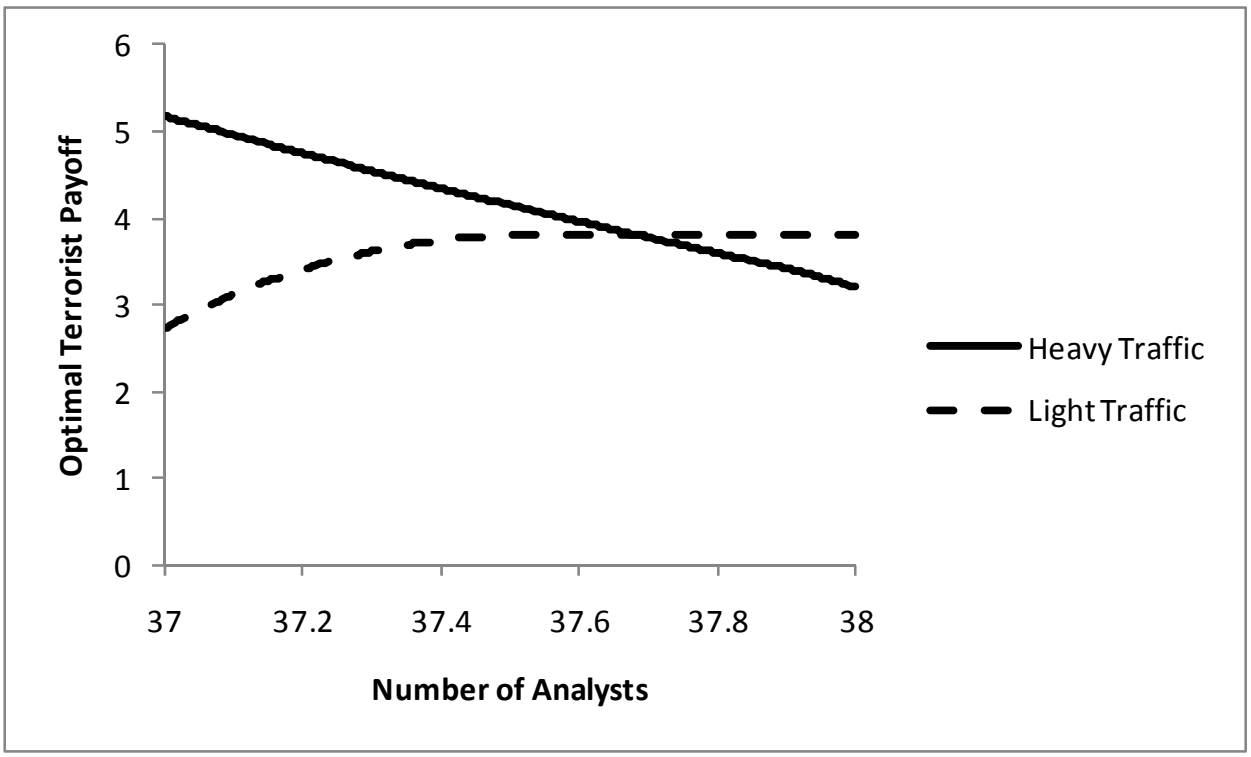

Figure 1: Equilibrium when the government moves first.

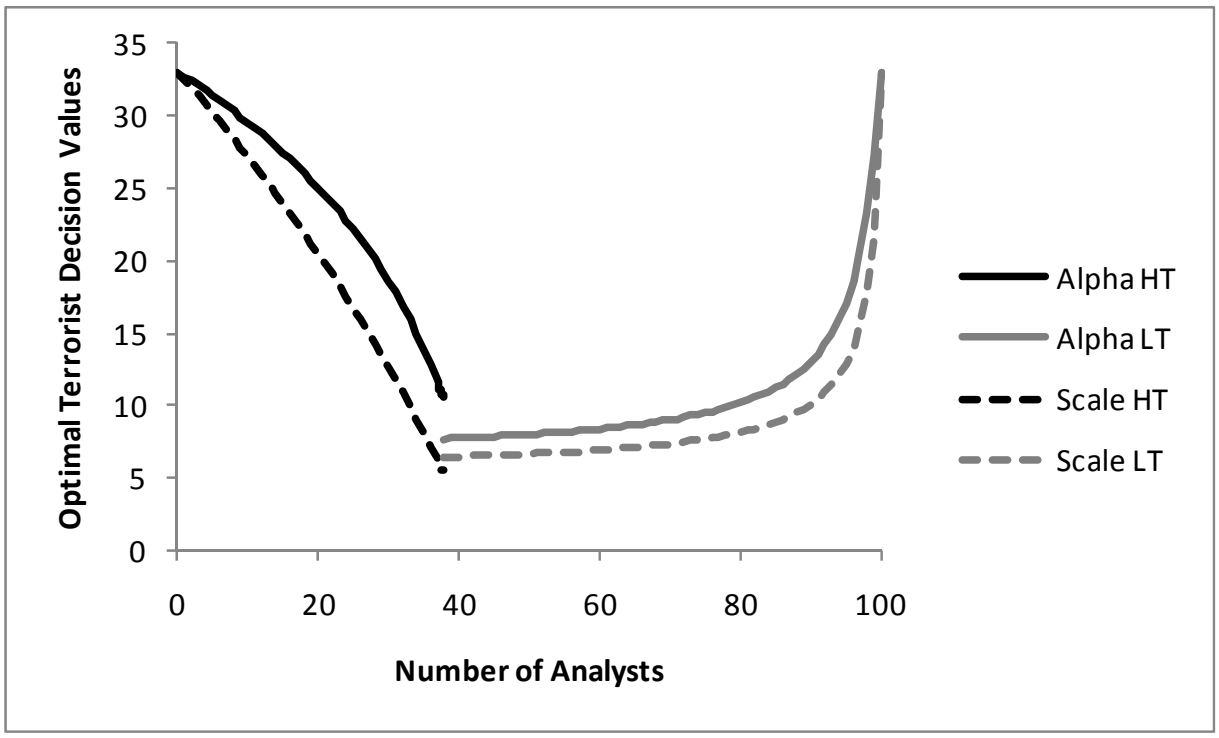

Figure 2: Optimal terrorist decisions when the government moves first.

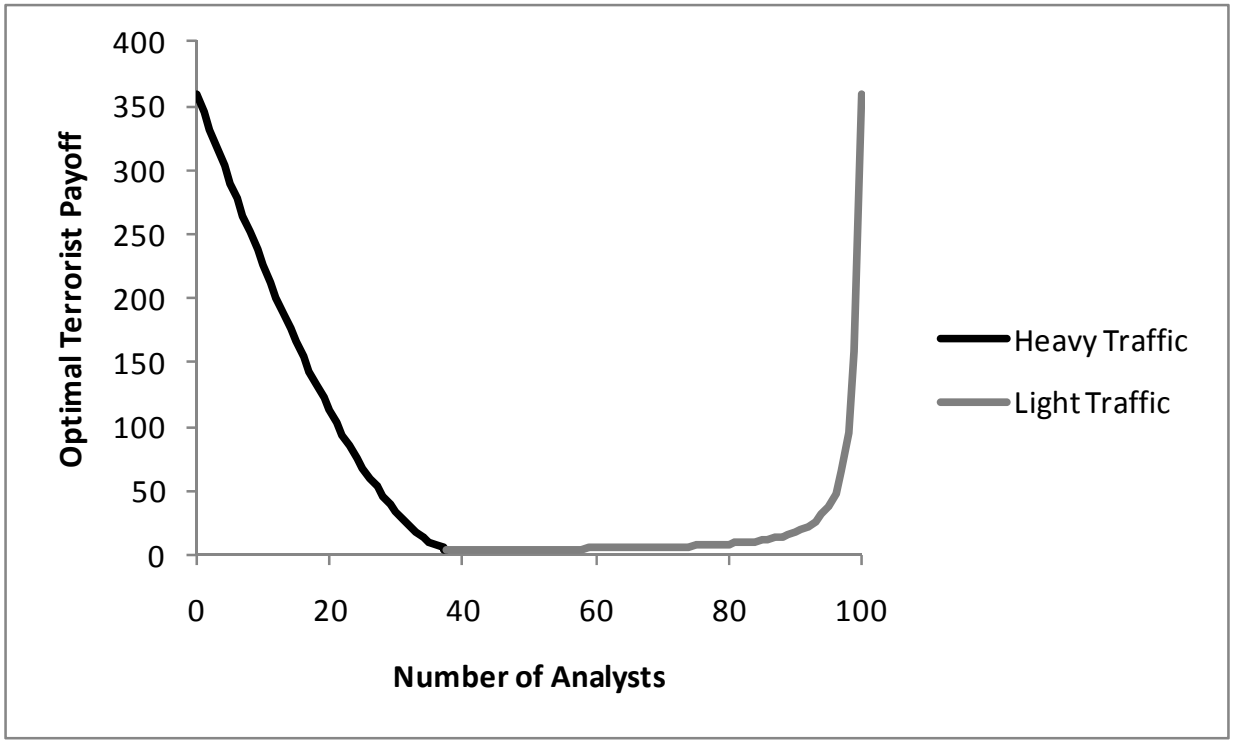

Figure 3: Optimal terrorist payoff when the government moves first.

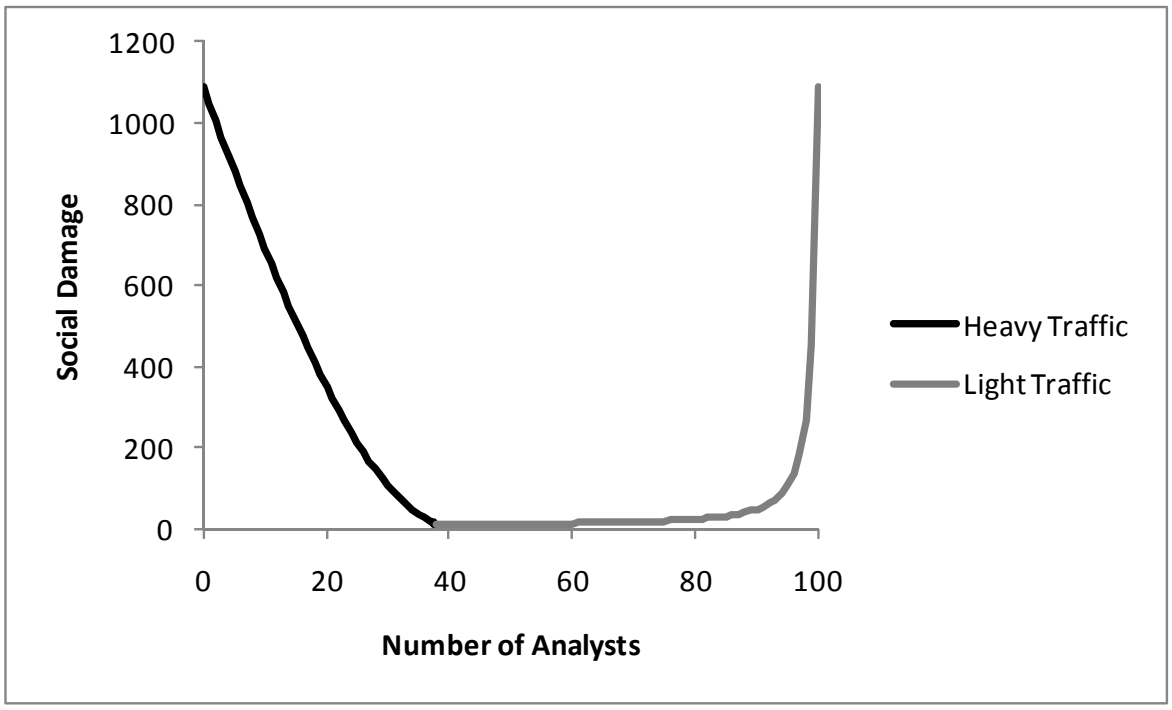

Figure 4: Social damage when the government moves first.