Center for Technology Innovation at BROOKINGS

December 2013

Cyber-enabled Competitive Data Theft:
A Framework for Modeling Long-Run Cybersecurity Consequences

Allan A. Friedman, Austen Mack-Crane, and Ross A. Hammond

INTRODUCTION

Cybersecurity has become a pressing policy issue, and has drawn the attention of the national security community. Yet there is an emerging consensus among experts that one of the largest policy problems faced in cyberspace may be not a question of military threats in a new domain, but the massive exfiltration of competitive information from American companies. Economic espionage has existed at least since the industrial revolution, but the scope of modern cyber-enabled competitive data theft may be unprecedented.

Much of the conversation surrounding the impact of cyber-enabled data theft has focused on how much theft is occurring today and how much this theft costs our economy today. Since data on the former (the level of theft) is extremely limited and almost certainly incomplete, efforts to estimate the latter (the present cost of theft) have suffered from both limited data and analytical approach, leading to widely varying estimates. Our focus in this paper is instead on *long-term* consequences of cybertheft for innovative sectors of activity that are at the core of US economic success. We conceive of the problem as one of diminished growth, rather than purloined assets. We explore the long-run implications of a world with no more (or with selectively fewer) digital secrets, examining which sectors or industries will be hurt the most or remain resilient, and which policies or technologies might be priorities for limiting economic harm in the future.

We begin by developing a framework to unpack the concept of "cyber-enabled competitive data theft" (CCDT), which comprises many different dynamic pathways. The type of data stolen is important: even files typically seen as mundane, such as email archives, could be of great value to an attacker. The right emails can reveal a bidding strategy for a billion-dollar deal, for example. We also consider how different

Allan A. Friedman, PhD is a fellow in Governance Studies, and Research Director of the Center for Technology Innovation at the Brookings Institution. His work focuses on information technology policy, with a particular expertise on the economics of cybersecurity. He is the co-author, with Peter Singer, of Oxford Press's forthcoming book *Cybersecurity and Cyberwar: What Everyone Needs to Know*.

Austen Mack-Crane is a research assistant in the Center on Social Dynamics and Policy at the Brookings Institution. He has experience designing and implementing numerical and agent-based models in areas such as public health, natural resource management, and game theory.

Ross A. Hammond, PhD is a Senior Fellow in Economic Studies at the Brookings Institution, where he is Director of the Center on Social Dynamics and Policy. His primary area of expertise is modeling complex dynamics in economic, social, and public health systems using mathematical and computational methods from complexity science. His current research topics include obesity etiology and prevention, food systems, tobacco control, behavioral epidemiology, crime, corruption, segregation, and decision-making.

protection "regimes" (investments in particular forms of cybersecurity) map onto what types of information are or are not effectively protected. We detail the types of data that any firm might use to create value that are also of interest to attackers. These classes of information can be mapped to industries and sectors based on how attackers use strategic information. We then explicitly catalogue how firms suffer direct, first-order harms from data theft. In the model, we instantiate industry-specific patterns of information use related harms from theft drawn from extensive case studies, interviews, and the published literature. We then model expected long run shifts in the distribution of production and investment in innovative activity resulting from any particular pattern of harms.

CATEGORIZING AND MAPPING DATA THEFT

To model the economic effects of data theft, we need to characterize the data, and the ways in which theft of this data harms the victim. In public discussions of this problem, many assume "the data" refers only to advanced technology, which would allow the exploiter to replicate the owner's products and compete in global markets in high value industries. While this is undoubtedly a major concern, competitive data theft spans beyond what we might think of as the drivers of high-end innovation. We divide competitive data into two broad categories: *proprietary technology data*, and *tactical data*. Proprietary technology data, under a broad definition of technology, can be anything that supports the creation of a good or service. Tactical data, on the other hand, supports company decisions, from pricing information to long-term strategies. These two categories cover all of the pathways through which market actors can create value from data, while not including other forms of intellectual property that are the products themselves, such as copyrighted works.

Proprietary technology data directly informs the product, and any theft of such data might allow a competitor to introduce a similar product, or improve their own product. We characterize this type of data as a spectrum. At one end are formulas, blueprints and other data that directly inform the creation of the product. At the other end is the supporting knowledge that informs the generation of future products, such as research or market measurement. We characterize the middle of this spectrum as 'process' data, or knowledge that relates to specific ongoing processes. This might have immediate use, such as a catalytic process that improves the quality of a manufacturing process, or be harder to adopt, such as the quality control processes of a manufacturing plant. We make this spectrum discrete using the three buckets above as distinct categories for different types of data.

Tactical data can be broadly defined as any information that informs a company's commercial decisions. Of course, this category includes practically all non-product related data in a firm, but that's precisely the point: it can all be valuable. Tactical data is often time-sensitive, which can help and hinder. On one hand, that limits its potential value to would-be attackers. On the other hand, in the right (or wrong) circumstances, use of this data can be seriously disruptive.

Major sales deals, supplier contracts, public auctions or mergers & acquisitions can turn on a few key pieces of data. Shortly after Coca-Cola was attacked, the Chinese government rejected their bid to purchase a major Chinese soft drink bottler, derailing a deal worth billions. Unfortunately, such incidents are hard to capture, as failed deals can have many potential causes. Our evidence set also captures a handful of cases where the thief stole tactical data such as sales contacts, customer proposals, pricing information or corporate strategies. Estimating the relative value of this for a given sector requires understanding a series of components, including the concentration of the customer base, the nature of the sales process and market for raw materials, and the relative churn of market actors, M&As, etc. Other experts talked of the importance of business process data, such as logistics. We focus on four main categories of data: sales data and plans, supplier and upstream data, information about bidding and strategic planning, and structural data about the long term plans of the enterprise.

Harms

Theft of data doesn't have to be bad for the victim. Even if the information ends up in the hands of a competitor, the competitor must be able to take advantage of it in a fashion that harms the owner. How might a firm be harmed? We have identified six different vectors of harm. The first four fall under the heading of 'lost sales,' as lost revenue is the most immediate and obvious harm to a competitive enterprise. Sales can be lost through different mechanisms, however. A competitor *can improve the quality* of a product, entering a market they otherwise would not have been able to enter. They could also be able to *reduce the costs* of a product whose capacity was already there, undercutting the rightful owner of the data. For example, while there are many ways of making certain forms of titanium dioxide, doing so cheaply and efficiently is a closely guarded process. Alternatively, competitors might be able to *poach sales* through knowledge gained from tactical data, such as sales contacts or bidding information. Finally, the victim could be harmed if theft of the information *reduced the value* of their product. The data stolen from RSA in 2011, for example, was used to compromise their SecureID product, reducing its value to customers and imposing replacement costs on parent company EMC. In this case, the competitor may not necessarily gain from the lost sales, but the product is less attractive to customers. Beyond lost sales, the competitor could disrupt the victims business by securing *preferential access* to a key input or upstream supplier. More generally, they could disrupt the victim by *interfering in their strategic plans*, such as by blocking a merger or making it more expensive.

It's important to acknowledge the limited scope of harms considered. We only consider first order harms at this stage, rather than trying to anticipate other indirect or collateral damage from cyber attacks. Second, we only look at direct economic harms. We do not address reputational issues where the news of a cyber attack is itself damaging. Finally, we only focus on cyber attacks on the enterprise. This leaves aside other intellectual property questions, such as unlicensed patent use, or illegal copying of digital goods already on the market.

Sectors

Our approach is built on the idea that different types of data are used differently by adversaries to gain an advantage, which in turn can inflict different damages on to the victim. Rather than characterize individual companies—an infinitely heterogeneous set—we focus on specific industries. As there is still vast heterogeneity across firms inside a given sector, we make a set of assumptions about what the firms have in common with each other and what separates sectors from each other. Different sectors face different types of competition, which shapes how potential thieves may be able to use stolen ideas in a competitive environment. Like the old joke about a physicist modeling a farm, instead of assuming a spherical cow, we assume common elements of data dependency and competitive risks.

Below we briefly summarize how each of the five sectors studied use data, and the risks they face from competitors using foreign data. These assumptions are built on industry data, academic studies, and interviews with experienced industry participants and trade representatives. We scoured Department of Justice press releases, government reports and congressional testimony, and the academic literature to build a dataset of known cases of competitive data theft with established estimates of harms. (In this phase, we studied a full range of competitive data theft vectors, with a malicious insider being the most common vector of data exfiltration.) We focused on reports with some metric of harm for two reasons. First, it offered insight into the different mechanisms of harm estimation, requiring the victim to actually suffer some type of harm. Second, this offered empirical proof of the range of harms, ranging from no harms suffered to losses of billions of dollars. In parallel, we reviewed the organizational science literature to understand the different theories of how information can be used to create value, in an effort to identify how compromise of that information might interfere with that process. We also use industry data from a range of sources, including government data, to characterize differences between sectors. Finally, we spoke with over a dozen experts from industry to get specific perspectives about how their businesses use data to gain competitive advantages, and the perceived risks of data theft, as well as cybersecurity experts with specific insight into the risks of data theft.

First, we look at the *chemical sector* as an archetypical knowledge-driven sector. It is an industrial sector, but one with both secret formulas that might be stolen, as well as secret processes and other advantages for better or cheaper production. Competitors could thus gain advantages by either improving the quality of their own goods, or introducing cheaper goods without the need to cover research costs. We have identified examples of both, but cost competition is more common. Sales tend to be a bit longer term in this sector, so poaching is less of a risk than the potential of losing access to inputs. The chemical sector is not particularly dynamic, so the risks of strategic disruption are below average.

In contrast, the *pharmaceutical sector* is very dynamic, as larger firms adopt a model of buying leaner, single-purpose research-based startups. In the US, drugs compete in narrow niches, so poaching is less of a risk, and rare ingredients are less common. On the other hand, the pharmaceutical sector has been plagued by concerns about intellectual property, particularly in terms of patent violations by third party manufacturers. A third party can obtain public information about drug formulas of successful drugs, after the expensive research and testing process has been done, mass-producing them and denying the innovator of revenue to recoup its expenses. However, it is exactly because this data can be obtained through other means that cyber theft poses less of a risk. This is not to say there is no concern from data stolen directly from the company—a handful of cases demonstrate that large pharmaceutical companies should worry about it a little, but the main threat is strategic disruption.

We include the *financial sector* to examine a sector that is data-intensive, but often ignored in discussions of technology. We do not include the retail side of the sector, and instead model an enterprise whose competitive advantage is in providing returns to capital. This includes firms that sell solutions, or take large market positions based on data. Because the actual data that is used by particular firms is often available to other market players, the competitive advantage relies in innovative application of this data. Given the fast pace and large returns in this sector, financial firms can be vulnerable to competition from theft of an idea, either in complete form or its basic building blocks. Competition is particularly fierce in terms of stealing customers, or losing unique access to a source of data or technical advantage. A financial firm is also uniquely vulnerable to the threat of its product losing value. Since an advantage also depends on uniqueness, any publicized knowledge of a tool or strategy can be used to counter that advantage, since markets generally exploit any arbitrage opportunity down to zero profitability.

Consumer electronics covers a large space of products, so we focus on mobile phones. This sector is characterized by rapid innovation but relatively short windows of profitability before the next latest-and-greatest product arrives on the scene. There is strong price competition, since last year's brand new feature becomes this year's standard, mid-market product. As such, competitors can leverage technical details to compete on cost and, to a lesser extent, quality. There has been notable churn in the mobile sector, but it is mostly been driven by new companies supplanting older companies; the mergers and acquisitions are a secondary consequence, often driven by the need to hold large patent portfolios. The mobile space is particularly vulnerable to a specific type of value loss: future product or pricing information can be vital in helping to gain a strategic advantage in the market place if a competitor can time the market or consumer reaction just right.

In contrast, the *semiconductor industry* is less surprisingly less vulnerable to many types of cyber attacks. While driven by innovation and scientific knowledge, the scale of high end

chip manufacturing requires vast capital investment. The market has begun to vertically fragment, with foundries just focusing on making chips and fab-less designers contributing the key intellectual property. Another chief reasons we believe that the semiconductor industry is not quite as vulnerable to CCDT-related loss is the relative immaturity of the Chinese semiconductor industry, after over a decade of massive investment. That said, customers can still be poached, and the fight to have ones chips in key products is cutthroat.

For each of these sectors, we define a two-dimensional mapping of how exploitation of the specific types of data mentioned above could be used to inflict a specific type of harm. For example, both the financial sector and the consumer electronics sector can lose the value of their products if certain data are stolen. In the financial sector, that data pertains directly to the product, and the process to a lesser extent. For the consumer electronics sector, it is the tactical data relating to sales and structural plans that can allow a competitor to undercut a product launch. This process has an inherent amount of subjectivity, but we worked hard to inform each assumption with evidence from the database, or particular insights from interviews, industry data and analysis, and our expert interviews. We defined the harms on a scale of 1-10. Initially, each value was set to an average of 5. We then identified instances where it was highly unlikely that stolen data could lead to a specific harm. For example, tactical data about a supplier could potentially help a competitor in the chemical field improve the quality of their offering if they could glean insights into the manufacturing process, while this is unlikely for the pharmaceutical industry. Following this, we defined the potential for harms in a relative fashion, using the numeric scale to note when the potential harms were higher or lower than an 'average' victim of data theft. These specific risks from each data type were then aggregated to produce a single number on the same scale for each potential harm for each industry, collapsing three dimensions into two. These values are presented in Table 1.

TABLE 1

ESTIMATES TO THE RELATIVE HARMS OF EACH SECTOR, FROM THEFT, AGGREGATED FROM ALL TYPES OF DATA

TYPE OF HARM	CHEMICAL	PHARMA	FINANCIAL	CONSUMER	SEMI
COST	8	2	7	8	6
QUALITY	7	6	3	7	5
POACHED	5	5	7	3	4
DISRUPTION	5	8	7	5	4
VALUE	0	0	8	4	0
INPUTS	6	2	6	6	3

Building the model

As discussed above, one goal of this study is to understand the complex nature of data theft, and how different technical and policy options might affect long term outcomes. In our model, sectors experience harms from data theft that reduce their subsequent growth rate. Each sector is treated as one of a set of options for an investment community. The theft of competitive data makes that sector less attractive to investors, and the sector further suffers from the loss of capital.

Since the model is temporal, we had to determine the timing of harm application to firms. Here, the focus on economic growth offers large advantages. Recall that we reject an approach where stolen data is like a stolen asset, reducing the value of a company by a fixed cost. Instead, a competitor can use stolen data to grow at the expense of the victim, so an incident of CCDT harms the future growth potential of the victim. The damage to this growth rate is permanent, although we do capture continued growth through innovation post-theft as long as the rate is not reduced to zero. To minimize unnecessary assumptions, we compare data theft across sectors at a constant time. If we had information of differential rates of theft, we might be able to introduce variable or stochastic timing, but one clear lesson from previous studies of large technical cyber operations is that they appear to hit sectors equally.

For the initial model, we wanted as simple a model as possible to minimize complexity and limit the potential for artifacts from the modeling process to drive results. To that end, we collapsed the different types of harms into a single value. Again, the goal is to capture the

dynamics of CCDT, so these values are built on a process that captures the relative harms. This requires further simplifying assumptions. To derive the harms from the theft of all data, we averaged across all values above, to come up with a value between one and ten, with ten being conceived as a lethal or near-lethal hit to an organization. We then reviewed these values again in light of our prior research and adjusted values accordingly. These values are the first column in Table 2, representing the theft of all data.

TABLE 2

SIMPLIFIED SINGLE-DIMENSIONAL HARMS FROM DATA THEFT FOR EACH OF THE FIVE SCENARIOS STUDIED

	SCENARIO				
SECTOR	ALL DATA THEFT	TECH SECRETS	TACTICAL DATA	INT: LEGAL	INT: TECH
CHEM	5	3.5	1.5	5	4
PHARMA	3.5	1	2.5	3.5	3
FINANCIAL	6	2.5	3.5	6	3
CONSUMER	5.5	3	2.5	4	5.5
SEMI	3.5	2.5	1	1.5	3

We explore five different scenarios of data theft. In addition to considering the harms of all data theft, we differentiate between the different types of data stolen to demonstrate the importance of considering them separately. We look at 'classic' intellectual property theft, where the adversary just seeks product-related *technical secrets*. We compare that with theft of just *tactical data*. We then consider two potential solutions to CCDT. In the first instance, we explore how a more active international intellectual property regime might mitigate harms. If firms or countries could seek legal redress through some legal protection treaty or some other intellectual property forum, it might deter provable instances of data theft. Alternatively, firms could invest in some moderate data protection. This would not eliminate all data theft, but would reduce it, particularly for data that is concentrated and accessed by a small set of users. Securing large, distributed data sets that are accessed by many different actors across multiple platforms is much harder, so we focus on the more feasible option. These two scenarios are explained further below.

AN INITIAL MODEL OF SECTORAL CONSEQUENCES

We present a model of sector growth that represents productivity growth, investment dynamics, and exogenous harms from cyber-enabled competitive data theft. The model economy includes the five sectors described above, which interact indirectly through the allocation of investment. Harms cause permanent reductions in sector growth, and are applied to all sectors at the same time, early in the run.

The basic model assumes that each sector is identical except for the different consequences of data theft. Each has a growth function (shown below) that determines how it will convert the capital from the investor pool into productive output. The model reallocates capital based on current productivity in each time period.

The model draws the harms information into a data structure holding the harms of each type (and for each information type) to be applied to each sector at each timestep. After initialization, at each time step the model solves for the allocation of the money supply between the five sectors. Given that allocation and the harms for the time period, it asks each sector to calculate its growth for that period. This is repeated for T periods, and data are collected on sector size and investment by sector.

Growth Function

We use a traditional Cobb-Douglas production function to model output by sector:

$$Y_i = A_i K_i^{\alpha} L_i^{1-\alpha} \quad [1]$$

In eqn [1], K denotes capital stock, L denotes labor, and A represents productivity, which grows at a certain rate r_A in each time period. Labor is not modeled dynamically, so it functions as a constant, and is included in order that output should have decreasing marginal returns to capital.

Dynamic Investment Allocation Algorithm

We model investment as an additive increase to firms' capital stocks:

$$Y_i = A_i (K_i + I_i)^{\alpha} L_i^{1-\alpha} \quad [2]$$

and assume that capital is compensated at the rate of its marginal productivity:

$$\frac{\partial Y_i}{\partial I_i} = \alpha A_i L_i^{1-\alpha} (K_i + I_i)^{\alpha-1} \quad [3]$$

The equilibrium investment allocation is found by computing the investment allocation for which marginal returns to investment are equal for all sectors, or $\frac{\partial Y_i}{\partial I_i} = \frac{\partial Y_j}{\partial I_j}\ \forall i, j \in \{1, \ldots, 5\}$ and

$\sum_i I_i = I$, where I is the money supply. We assume that the money supply grows at the same rate as overall output in the model economy.

Our computational model approaches this problem as a 5-dimensional maximization problem with a hyperplane budget constraint and non-negativity constraints, and it is solved using a constrained gradient ascent algorithm.

Incorporating Sectoral Harms

Harms are conceived of as affecting the rate of productivity growth, r_A, in a sector, and are independent of investment decisions in the current period, but affect future investment. Because we don't have theoretical justification for the magnitude of the harms in relation to sectors' productivity growth, we include a *free harm-scaling factor* (HSF) which we explore in our simulations. Growth for sector *i* at time *t* is computed by:

$$Y_{it} = A_{t-1} \cdot \left(1 + r_A\left(1 - \frac{h_{it}}{f}\right)\right) \cdot L^{(1-\alpha)}(K_{t-1} + I_t)^{\alpha} \qquad [4]$$

where h_{it} is the total harm experienced by sector i at time t, and f is the factor used to scale harm values in relation to r_A.

SIMULATIONS

We let initial output, initial productivity growth rate, α (the exponent of capital), and the initial capital-labor ratio be the same for all sectors. These assumptions are unrealistic, but we judge it difficult or impossible to construct estimates for these that would be more defensible than uniformity. α and the initial capital-labor ratio, in particular, were found in exploratory simulations to have very strong effects on sectors' growth trajectories, and if it were feasible in future work to empirically estimate a growth function for each sector, this would seem to be a fruitful extension.

The parameters for our main results are set as follows:

$$\alpha = .5$$
$$\frac{K_0}{L_0} = 1$$
$$r_A = .025$$
$$T = 100$$
$$I_0 = 100$$
$$Y_0 = 100$$
$$f \in \{1, 2, \ldots, 10\}$$

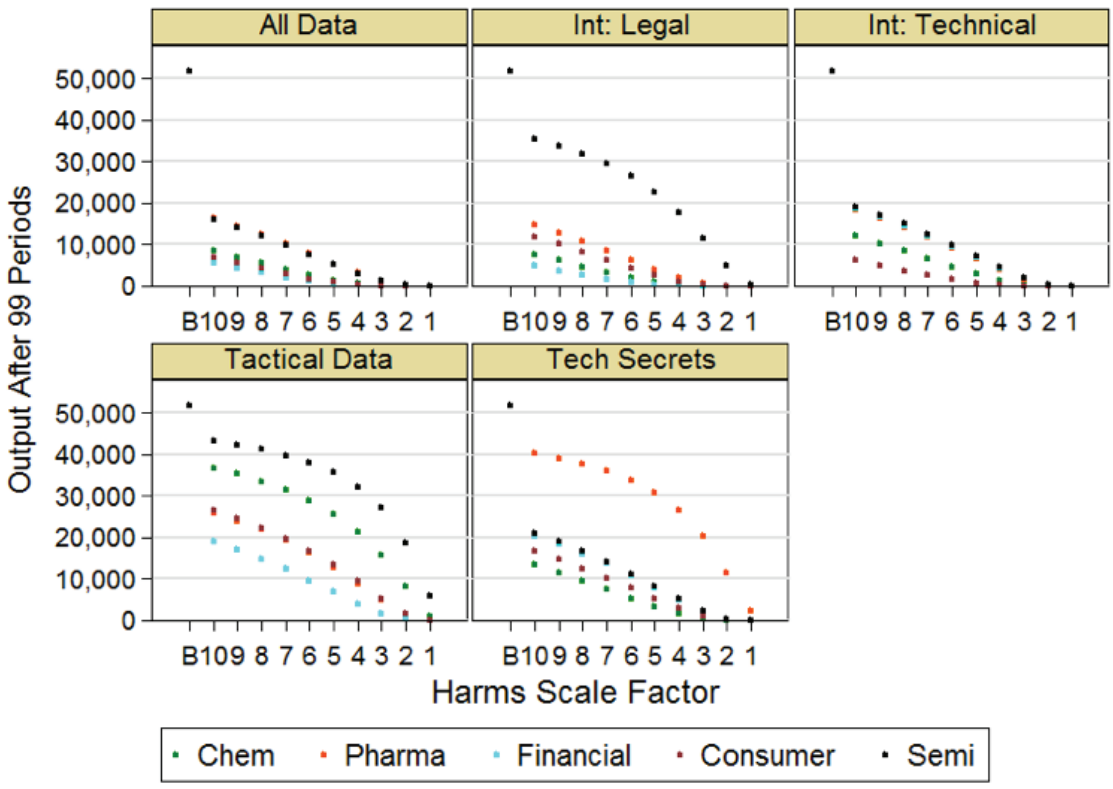
MODEL RESULTS

Our simulations collect data on output and investment over time for each sector, and allow us to compare sectors' long-term performance under different scenarios. Because sectors are initialized with the same output level, end-of-simulation output is an appropriate metric for relative performance under each scenario.

Figure 1 gives a summary of the sector output results across different harm scaling factors and under different harm scenarios (each in a separate panel), using output from the final time period of the simulation. Each data point represents the final output level (y-axis) of a particular industry (color coded) for the particular HSF parameter chosen (x-axis). In all of the panels, output levels are lower as the scaling factor decreases, due to the fact that HSF divides the harm magnitude, as in equation [4]—a higher HSF means less powerful application of the harms across all industries. Although all sectors start with the same output (at time zero) in the simulation, differences emerge in relative performance for many of the scenarios and HSF values, with clear "winners" and "losers", There exist scenarios and parameter settings for which all sectors decline to zero output (see Fig 1 "All Data, HSF=1"), others where some sectors flourish and some die out (see "Tactical Data, HSF=2"), and many instances where all sectors grow at different rates (see "Tech Secrets, HSF=6). Even when all sectors are growing, those with different productivity growth rates still diverge in absolute terms due to their differing exponential growth paths.

FIGURE 1

THE HARM SCALING FACTOR (HSF) SETS THE RELATIVE DAMAGE FROM DATA THEFT. THE OUTPUT OF EACH SECTOR VARIES ACROSS SCENARIOS, INDEPENDENT OF THE IMPACT OF HSF. (THE VALUE "B" ON THE HORIZONTAL AXIS DENOTES THE BASELINE CASE WITH NO DATA THEFT.)

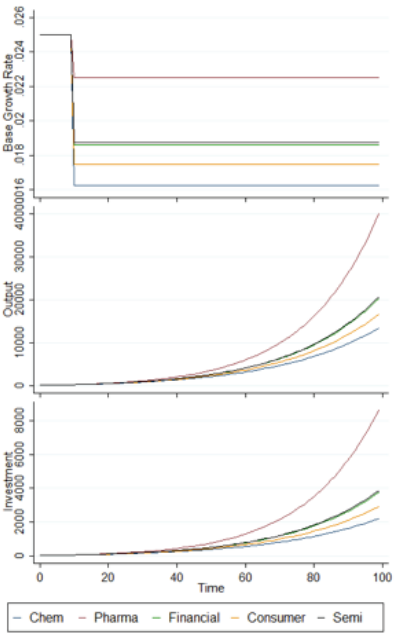


Given the input evidence and assumptions described above, these results yield insight into the central question of the paper—the potential long-run cross-sectoral impact of different types of CCDT.

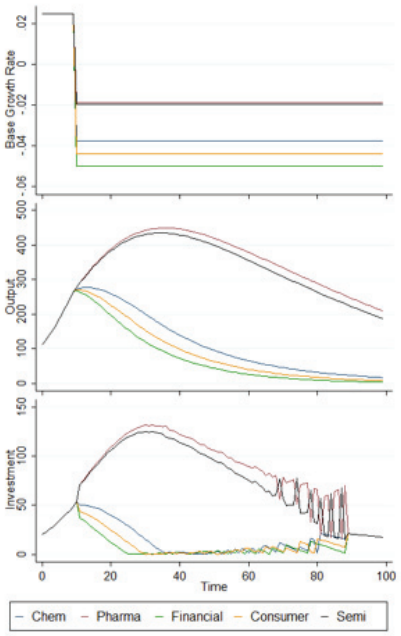
Figure 2 gives examples of more detailed outputs from the model, showing as a time series the dynamics of productivity growth rate (illustrating the effect of harms), output (with the final time point reflected in Fig 1 above), and investment for all sectors, focusing on one harm scenario and one factor at a time. We include three panels, including examples where all sectors grow (2A), where all die out (2B), and where outcomes are mixed (2C).

FIGURE 2

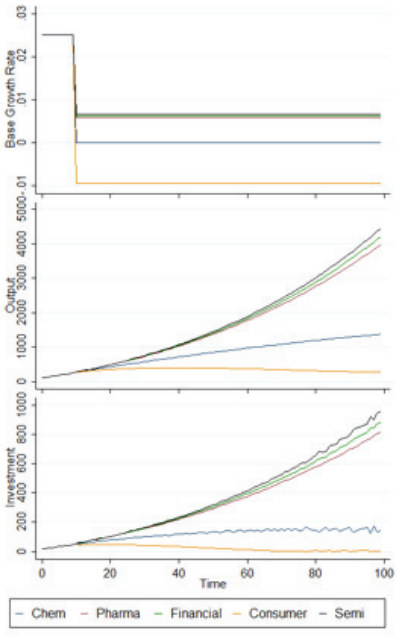
DIFFERENT TYPES OF THEFT LEAD TO DIFFERENT LONG TERM EQUILIBRIA



[A]



[B]



[C]

Example plots for (A) Tech Secrets scenario, HSF = 10; (B) All Data scenario, HSF = 2; (C) Technical Intervention scenario, HSF = 4. Plot (A) gives characteristic curves for the case where all sectors grow, but at different rates; plot (B) demonstrates the case where all sectors die off; and (C) shows a case where some sectors grow, one (Consumer) dies, and one (Chemical) stagnates. NB: the noise in investment in plot (B) is a germane feature of the investment algorithm that occurs when marginal returns to investment become very small.

Looking across all of the model results, several clear patterns emerge. In all cases, rates of growth and resulting output levels for the sectors are ordered as the inverse of the magnitude of harms applied (as expected, given that the input evidence drives model dynamics). Additionally, we can infer from our results (though not prove) that the sign of a sector's productivity growth rate after harms are applied determines whether it will grow, decline or (when productivity growth is zero) stagnate. Thus, in this basic model, growth and decline are independent of investment dynamics, though investment can affect their rate.

Understanding Interventions

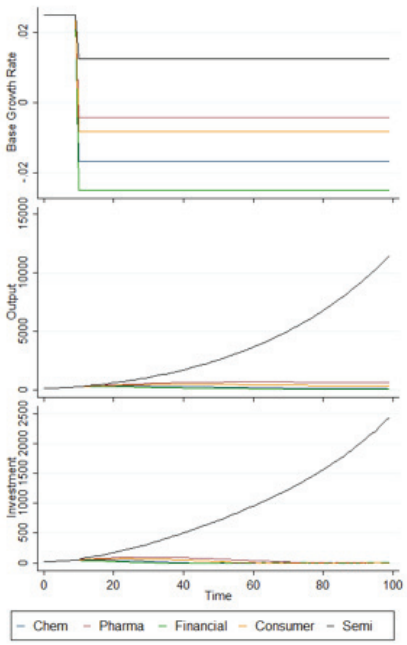
We next use our framework and model to consider potential changes to the status quo base model. We look at two different sets of stylized interventions. The first is a legal solution, imagining a hypothetical international enforcement of trade secret law. Under this scenario, we suppose that a victimized company that recognizes stolen data being used in a competitor's product might be able to enforce some penalty against the firm through a legal process. This process would help deter theft of formulas and other easily recognizable design components, but would not offer serious protection for the harms driven by theft of less identifiable technical data such as process secrets or research, and offer no protection for tactical data.

The second intervention is technical in nature, assuming some improved investment in data protection, reducing probability of data theft. This is not perfect protection. First, we assume that the probability is reduced, but not eliminated. The harms are at best halved. Second, data that is concentrated and used by a small set of individuals is much easier to protect than data that is spread across an organization and used by many different parties. As such, sectors like the semiconductor industry, with a global supply chain, will not gain as much protection as a smaller financial organization, who can invest in stronger access controls and data security without fundamentally reorganizing their entire firm. The model inputs for these scenarios are detailed in Table 2.

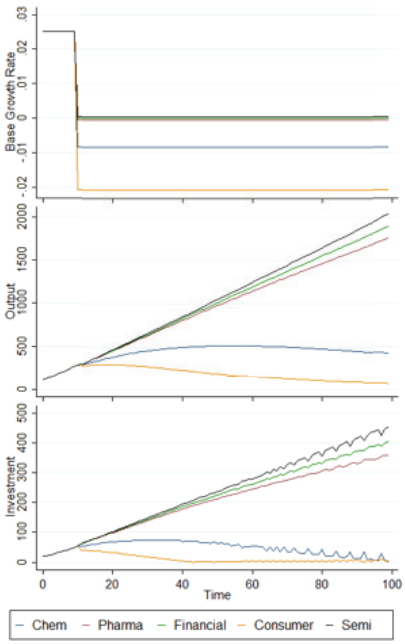
Figure 3 presents the results of these two interventions with a common HSF (to allow effective comparison). In Figure 3[a], we see that for the specific HSF chosen, the protections afforded by legal protection primarily benefits the semiconductor industry, where relatively straightforward identification of both stolen designs and stolen advanced processes through basic engineering forensics is possible. While the consumer electronics sector also derives some benefit, it is not enough to change the sector-comparative output of the model. For the chemical and pharmaceutical industries, legal protections might help, but they are already part of the status quo for pure stolen formulas, so there is no value added from our hypotheticals. In the financial sector, ideas are copied all the time, so a legal protection regime would not make any difference.

FIGURE 3

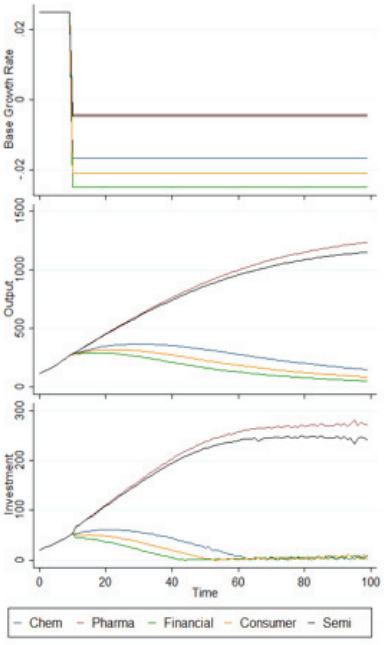
COMPARISON OF (A) LEGAL AND (B) TECHNOLOGICAL INTERVENTIONS, INCLUDING (C) THE ALL DATA SCENARIO AS BASELINE, ALL WITH HSF = 3.



[A]



[B]



[C]

In our scenario modeling data protection (Figure 3[b]), we see a different outcome (again contingent on the choice of HSF). The semiconductor sector is still high-performing, but now we can see that the pharmaceutical and financial sectors are expected to flourish. If a drug company can protect a relatively small amount of data on its long term plans, it will not be nearly as vulnerable to harms from market disruption. The financial sector is concentrated, as discussed above, especially compared to the massive research and development facilities of, say, the chemical sector.

These simulations help demonstrate the power of this model by testing various hypothetical policy and market solutions to the problem of CCDT.

Sensitivity Analysis

Following best practices for dynamic modeling, we consider two extensions of the model presented to test the sensitivity of our results to our assumptions. First, we include a sixth, unharmed sector in the economy, to represent the collection of industries less susceptible to data theft. This creates greater competition for investment funds, but also greater money supply growth than in the five-sector case. This alternative model does not change the qualitative facts just described, though it does cause growing sectors to grow more slowly than they otherwise would.

A further extension we explore, building from the 6-sector model, is to allow capital to depreciate. One clear change that this causes is that sectors with zero productivity growth now decline (though they may eventually asymptote instead of dying off entirely). We also see higher investment in declining sectors than we did in the basic model. Otherwise the qualitative observations above appear still to hold.

DISCUSSION & CONCLUSION

We have presented what we believe is the first economic framework and model to understand the long-run impact of competitive data theft on an economy by taking into account the actual mechanisms and pathways by which theft harms the victims. Data theft was understood by looking at the harms to future growth and productivity. By focusing on differences between industrial sectors and how they use data, we show that data theft will not have a homogenous, undifferentiated impact on the economy. Instead, the harms from any particular type of data theft are likely to impact sectors unequally in many circumstances, triggering potential long-run shifts in the distribution of productive economic activity in the US. We believe that the general approach we describe is an important new direction for research on cyber-theft of competitive data—enabling a more nuanced and mechanism-based evaluation of diverse types of harms through time. This requires sector-specific evidence, consideration of distinct uses of data in different forms, and a dynamic modeling approach.

Given the limited evidence presently available, our harms matrix, our model design, and our quantitative conclusions are necessarily preliminary and exploratory. Nevertheless, the initial use of the framework presented here serves as a "proof of concept" for the approach, and our initial results suggest five important conclusions.

First, as our results in Fig 1 and 2 demonstrate, the three dimensions along which our framework differentiates CCDT can all be important to model outcomes. In some cases, sector matters, in others the type of data stolen matters, and in others protection regime matters.

Second, by seeing stolen data from a business process perspective, rather than a lost asset, we were able to understand the problem in a longer time frame. This not only avoids the challenges of short term analysis and gives us the context of equilibria, it is more extensible in a policy analysis. For example, we could extend the model to test the value of transparency. A firm might, either through ignorance or deception, fail to disclose data theft. If the lost capacity is not immediately evident, we can model the impact on investors of this temporary ignorance.

Third, these simulations demonstrate that different interventions will have different effects. Not only is there no 'silver bullet,' but some sectors will benefit from solutions that may offer no help to others. This has clear ramifications as US policy-makers tackle this problem at a national and international scale. Moreover, it is also relevant to technical experts and information security vendors. Our research process helped us appreciate many ways data is stored and used, and the differences in cost and ease of protecting different data architectures. This informed our decision to avoid assumptions of perfect data protection in our data protection scenario. It also reinforces the notion that technical protection against CCDT requires not only technical protections, but organizational adjustments to make data protection investments easier and more effective.

Fourth, our framework introduces a new way of thinking about cybersecurity that does not easily map onto existing theoretical structures or evidence. The modeling process revealed the need for further theoretical work to properly integrate the diversity of impacts the framework identifies into a model of growth. In addition, we have identified clear needs for a richer evidence base to support appropriate comparisons of harms with existing growth parameters.

To avoid extending beyond the supportable base of empirical information, our initial dynamic model makes a number of simplistic assumptions that might repay further investigation. We treat sectors as uniform except in their susceptibility to data theft, ignoring important differences in size, market capitalization, productivity growth, and other factors. We don't dynamically model responses by firms, such as investments in data protection or legal action, or adaptation on the part of attackers.

Finally, this basic model is not only extensible, but can help us understand a range of critical cybersecurity policy problems. A particularly promising extension of the model would be to divide each sector into two groups: defenders and self-insurers. The defending firms spend some of their fixed capital in a one-time investment, but are less vulnerable to attacks. The remainder of the sector chooses to use their capital for growth, as before. We can then characterize the impact of such investment decisions, as well as explore the relative impact of different costs to achieve a given level of security.

The framework and analysis presented in this paper represent a first step toward an ultimate goal of not only understanding the underlying mechanics of cyber-enabled competitive data theft, but also understanding the solution in an economic context. The Obama Administration has stressed the importance of incentives and market forces in driving investment in securing critical infrastructure. Our research begins the process of expanding this approach to securing America's competitive data, and suggests an important new direction for the study of this pressing problem.

Governance Studies
The Brookings Institution
1775 Massachusetts Ave., NW
Washington, DC 20036
Tel: 202.797.6090
Fax: 202.797.6144
brookings.edu/governance.aspx

Editor
Christine Jacobs
Beth Stone

Production & Layout
Beth Stone
Camden Richards

EMAIL YOUR COMMENTS TO GSCOMMENTS@BROOKINGS.EDU

This paper is distributed in the expectation that it may elicit useful comments and is subject to subsequent revision. The views expressed in this piece are those of the authors and should not be attributed to the staff, officers or trustees of the Brookings Institution.