The Medium-Term Effects of Unavailability

January 2006

Avi Goldfarb [*]

Abstract

This paper develops a new method of identifying why current product unavailability has an effect on future brand choices. If the impact is solely due to changing preferences, then all competitors of the unavailable item gain proportionally to their share; if, however, there is lock-in, then the competitor that is chosen instead of the unavailable product should gain disproportionately more. Denial of service attacks at Yahoo, CNN, and Amazon show that unavailability has a medium-term impact. Lock-in drives 51% of the effect on Yahoo, but it dissipates much more quickly than the effect of changing preferences.

Keywords: product availability, stockouts, Internet, Denial of Service, switching costs

JEL classification: M3, L8

Product availability is a central concern in marketing channels. The purpose of marketing channels is to make “a product or service available for use or consumption” (Coughlan et al., 2001, pp. 3). However, even with recent improvements in supply chain management, products are often not available. Worldwide, retail stock-out rates are 8% (Corsten and Gruen, 2004). Product and service unavailability is also a frequent event for mail-order companies and online merchants (Fitzsimons, 2000).

The impact of product unavailability on profits depends on consumer reaction. If consumers merely delay purchase, then unavailability does not matter. Alternatively, if consumers permanently switch brands then unavailability is of vital importance. Effective channel management therefore requires an understanding of the impact of product unavailability on consumer behavior.

Broadly speaking, product unavailability has an immediate and a lasting effect on consumers. The immediate effect occurs when consumers cannot buy their brand of choice and therefore they either do not buy or they buy another brand. The lasting effect involves future choices. There are two broad reasons for the effect on future choices. First, the consumer’s opinion of the unavailable brand may have changed. For example, Swait and Erdem (2002) show that consumers assign a higher value to consistently available brands. Second, the consumer may have bought a different brand and have become locked-in to that rival brand. Consequently, the consumer may return to this rival brand in the future at the expense of the unavailable brand. This type of lock-in is well documented in consumer packaged goods. For example, Keane (1997) and Seetharaman, Ainslee, and Chintagunta (1999) document lock-in (also called “state dependence”, “loyalty”, and “switching costs”) in a number of categories. Optimal managerial response to unavailability depends on whether the impact on future choices is due to a decrease in the value gained from the unavailable product or due to lock-in to the rival product.

In this study, I measure both the immediate and future impact of one particular source of unavailability: Internet denial of service (DoS) attacks. I then identify whether the impact on future choices is due to a decrease in the value of a visit or due to lock-in. A DoS attack occurs when a hacker succeeds in shutting down a website, typically by programming thousands of computers to simultaneously request information from the website. This study focuses on a series of attacks that occurred in February 2000.

The results suggest that the impact on future choices is larger than the immediate impact. I estimate that Yahoo lost 2.22 million visits to its website during the attack itself. Future choices were also affected. The attacks cost Yahoo another estimated 7.56 million visits in the 53 days following the attacks. The effect on future choices dissipated over time. For this reason, throughout the paper I refer to the impact on future choices as the "medium-term" effect.[1] There is evidence of both a decrease in the value of visiting the unavailable website and an increase in lock-in at rival websites. Yahoo's rivals gained 2.78 million visits as a consequence of lock-in. This effect, however, was especially short-lived. In the days immediately following the Yahoo attack much of the effect was due to lock-in; 11 to 15 days later the remaining effect was almost entirely a change in preferences. Overall, 51% of the visits gained by Yahoo's rivals because of the attacks were due to lock-in. In contrast, only 13% of the visits gained by Amazon's rivals were due to lock-in.

DoS attacks provide an ideal setting to study the medium-term impact of unavailability on consumer behavior for three main reasons. First, and most importantly, rich data on online behavior make it easier to identify whether and how consumers are affected when a product or service is not available. Second, the exact timing was reported in the news. Third, website unavailability is a frequent problem for online marketers. Websites shut down for a number of reasons, including denial of service attacks by hackers, internal mistakes, local conditions (such as power outages in the server location), and Internet-level routing problems. Shutdowns are a sufficiently large problem that many websites have a disclaimer about availability in their terms of service. For example, Yahoo's (2005) Terms of Service make three separate references to availability. DoS attacks can be a particularly damaging source of product unavailability. Internet stocks fell substantially in the weeks after the DoS attacks of 2000 (most notably the attacked websites Amazon and EBay). In 2004, survey results suggested that 17% of websites experienced a DoS attack with an average immediate cost of $570,000 (Gordon et al, 2004).

I examine data on every website visited by 2,651 households from December 27, 1999 to March 31, 2000. The raw data contain 3,228,595 distinct website visits. This detail provides a natural experiment to separate the above two sources of a medium-term effect. The natural experiment arises because the data show whether a user visited a competing website to the attacked website. For example, the data show whether a user visited MSN.com during the Yahoo attack. This fact allows identification of lock-in (switching costs) at the rival website (i.e. MSN) separately from a change in preferences for Yahoo. If the impact is solely due to changing preferences then all competing websites should gain an amount proportional to their market shares (controlling for heterogeneity). If, on the other hand, there is a lock-in effect then the rival website that is visited during the attack should gain more than other competing websites. The benefits of lock-in only accrue to the website visited during the attack. Section 3 discusses the identification in more detail.

Through this natural experiment, this paper contributes to the existing literature on unavailability in three areas. First, it focuses on the impact of future choices. Most studies of brand unavailability have focused on the immediate impact (Jeuland, 1979; Campo, Gijsbrechts, and Nisol, 2000). Second, it uses real-world data rather than experimental data. Most existing studies focus on experimental data.[2] Third (and most importantly) this study separately identifies two broad reasons unavailability may affect choice in the medium-term: unavailability may reduce value (Swait and Erdem, 2002) or unavailability may induce consumers to try other products leading to lock-in. While other studies have examined the lasting effect (e.g. Fitzsimons, 2000; Bell and Fitzsimons, 1999), to my knowledge product unavailability has not previously been used to identify lock-in in any context.

The next section describes the data and provides a measure of the immediate impact of the attacks. Section 1 uses a difference-in-difference econometric methodology to show that DoS attacks negatively impacted Yahoo, CNN, and Amazon in the medium-term. Section 2 then shows that lock-in to the website visited during the attack played an important role for Yahoo and perhaps for CNN and Amazon. Section 3 estimates the model by segment and finds that the lock-in is largest for relatively infrequent users. Furthermore, the Yahoo attack mainly affected people who do not use Yahoo email.

This implies that simply emailing an apology to Yahoo's email users would have been a waste of resources. Section 4 builds a multinomial logit model of website choice and shows the qualitative results hold in this alternative econometric framework. Section 5 concludes.

1. Data and Immediate Impact

1.1 Description

The raw data set, courtesy of Plurimus Corporation, consists of every website visited by 2,651 households between December 27, 1999 and March 31, 2000 for a total of 3,228,595 observations. In addition, the data set contains the time of arrival at and departure from a website (to the second) and the number of bytes uploaded to the website. The richness of the data set allows for thorough examination of individual-level behavior.

The data have a number of limitations. First, they are collected at the household level rather than at the individual level. One individual could be online during the DoS attack and never online again during the sample. All other observations could belong to another individual. If this is the case, being online during the DoS attack will have no effect. Second, there are no at-work data. It is possible that members of the control group attempted to access a website from work during a DoS attack. Both of these limitations, however, will bias the results toward finding no effect for the attacks and I find an effect. Third, the data are not geographically representative. New York, Chicago, and Los Angeles are under-represented. Roughly half the sample comes from the Pittsburgh area. Another quarter is from North Caroline and another eighth from Tampa. Fourth, the data do not include AOL users. Finally, Plurimus collected the data from the ISPs. Their software recorded every page request sent through the ISP. This means that some cached pages may not show up in the data.

Table 1 shows that, despite these limitations, user behavior in the Plurimus data is similar to user behavior in the Nielsen/Netratings panel in the same time period. The top ten web properties are the same, although the order is slightly different. Furthermore, average time online per user is similar for the two groups.

This paper focuses on three of the seven well-publicized attacks that occurred in February 2000. Each of these attacks was conducted by the same teenaged hacker. Table 2 column (2) lists the exact timing of each attack. Table 2 also contains information that I use to help define which users tried to access the attacked website but could not.

Since the websites were inaccessible, I cannot determine whether a particular household tried to access the website under attack and therefore the treatment and control groups are not perfectly identified. Consequently, I estimate a probability for each household that it experienced the denial of service attack. First, if the household was not online during the attack or if the household had never visited the attacked website, then it is assigned zero probability of having experienced the attack. Column (5) of Table 2 shows the number of users in the data with a strictly positive probability of being in the treatment group. Second, household fixed effect probit regressions were run on the pre-attack sample to predict the probability that a given visit during the attack was to the attacked website. Section 1.2 provides more details on these regressions. To ensure robustness, I also show results where the treatment group is defined by users visiting websites in the same category, rather than online, at the time.

Based on these probit regressions, column (6) of Table 2 shows that only Yahoo, CNN, and Amazon have at least 30 users in the expected treatment group. There are not enough observations for EBay, ZDNet, Buy.com, and E*Trade to conduct meaningful statistical analysis. Therefore, the analysis that follows focuses only on Yahoo, CNN, and Amazon. Due to these sample size constraints, Yahoo in particular is emphasized throughout this paper.[3]

Columns (7) and (8) of Table 2 contain estimates of the immediate impact of the attacks. The estimates are based on the predicted probabilities of being in the treatment group and some assumptions on market size and revenue per visit. In particular, I assume a market size of 43.3 million online households based on Plurimus Corporation estimates. The revenue estimates assume 4 cents per visit. This value is based on revenue estimates from J. Walter Thompson Company for nine portals for January, February, and March 2000 combined with the visits data in this study. The revenue estimate is therefore based on data for Internet portals, and will be most reliable for Yahoo.

The three hour DoS attack on February 7, 2000 meant 2.2 million potential visitors could not access Yahoo. Yahoo promises its advertisers a certain number of page views and unique visitors each month (Yahoo Sales, 2000). I would have to make this up to advertisers. At four cents per visit, these foregone visits lead to an estimated direct loss of $88,854. The attack on CNN cost it 653,338 visitors and the attack on Amazon cost it 522,671 visitors. Given the frequency of DoS attacks and other website shutdowns, websites have a strong incentive to reduce shutdowns even without any medium-term effects.

A critical aspect of determining the effect of the attacks is defining the competitive set. These category definitions were set by Plurimus and are detailed in Table 3. There are 140 portals in the data, 182 news websites, and 366 shopping websites. Only websites that sold items that Amazon sold at the time are counted as shopping websites. Table 3 also lists the number of observations in the category, the number of users in the category, the market share of the attacked firm, and the market shares of the top ten competitors.

Table 4 provides descriptive statistics for the data used in this study. 'Media mentions' were constructed from the Lexis-Nexis Academic Universe database. The *# Media Mentions over past 15 days* variable is equal to the total number of days the attacked website is mentioned on network television news (ABC, CBS, or NBC) or in the New York Times over the previous 15 days. A media mention is also counted for local residents if a company is mentioned in the Pittsburgh Post-Gazette, the Tampa Tribune, the Dallas Observer, the Greensboro News and Record, or the Durham Herald-Sun. On average, Yahoo is mentioned 7 times, CNN is mentioned 1.5 times, and Amazon is mentioned 6.5 times. Bytes uploaded to the website proxies activity at the website. Bytes downloaded from the website gives similar results and is highly correlated. The different treatment group definitions in Table 4 are described in the next section.

1.2 Treatment Group Identification

This section describes the fixed effect probit regressions used to identify whether a household tried to access the attacked website during the attack. Table 5a presents the results of the regressions. Table 5b presents descriptive statistics of the predicted values.

The purpose of these probit regressions is to derive a household-level estimate for the probability of being in the treatment group.[4] In other words, the goal is to predict whether a given household tried to access the attacked website (Yahoo, CNN, or Amazon) over a particular time period. The regressions use all visits preceding the attacks rather than visits to competitors for two reasons. First, using the category undercounts the treatment group. During the attack, users who cannot visit their desired website may not visit any of its competitors. They are, however, likely to visit other websites during the online session. Second, the goal is not to predict category choice but to predict the choice of a particular website. Competitors are not relevant for determining the treatment group. The regressions therefore estimate the likelihood of visiting the attacked website instead of any other website. Only households that visited the attacked website at least once before the attacks are included in the regression. All other households are assumed to have zero probability of experiencing the attacks.

The covariates in the regression are household fixed effects, time of day (by hour) fixed effects, whether the attacked website was the previous website visited, the number of days that the attacked website was mentioned in the media over the preceding 15 days, the number of bytes uploaded to the attacked website on the previous visit, and a linear time trend. Fixed effects are used instead of random effects to improve fit and predictive ability. The coefficients of the regressions are presented in Table 5a.

The results of the probit regressions were used to predict visit probabilities during the attacks. In particular, the data on visits during the attack was combined with the coefficients of the probit regressions. If a household did not visit any websites during the attack, it has zero probability of being in the treatment group. If a household visited one website during the attack, the treatment probability equals the predicted value. If a household visited more than one website during the attack then the probability of being in the treatment group is the probability that the first visit was to the attacked website, plus the probability that the second was to the attacked given that the first was not, plus the probability that the third was to the attacked given that the first and second were not and so on.[5]

Tables 4 and 5b present descriptive statistics of this treatment group definition. Table 4 shows the means, standard deviations, minima and maxima. Table 5b presents details including distributions, comparisons between the treatment and control group, predicted number of users, and correlation coefficients with other treatment group definitions.

Part B of Table 5b compares treatment and control group characteristics. There are few differences in demographics measured at the census block level. Education and income levels are similar across groups. There are, however, differences in online behavior. Users with a positive probability of being in the treatment group email and chat more than users who are not in the treatment group and spend significantly more time online. The difference in time online is not surprising: users who spend more time online are more likely to be online during attacks. Table 9 and section 4 will show that the qualitative results hold if we restrict the sample to users who spend a great deal of time online. The predictive treatment group size is also similar to the actual visit propensities at the same time and day of the week (see Table 5b part C). The predicted number of users is calculated as the sum of all treatment group probabilities.

To ensure robustness, I use three other treatment group definitions to identify lock-in. While the above definition gives a good sense of whether the individual tried to visit the attacked website during the attack, determining the reasons behind the effect (section 3 below) involves understanding whether the visit to a rival website that was not attacked was out of character.[6] In this measure, labeled "relative to rival frequency" in the tables, I again use probit regressions to predict the likelihood of a visit to the rival website actually visited during the attack and subtract it from one. In particular, if all visits in that category before the attack are to that website then this has a value of zero; if the household has never visited the website before then this has a value of one; and there is a continuum of values in between. Rather than identifying how likely it was that the household experienced the attack, this measures whether the website visited instead was out of character.

Another measure defines the treatment group by the probability that a given household goes to the attacked website during the attack as its prior propensity to visit the website. This measure is labeled "relative to visit propensity" in the tables. Unlike the regressions, this estimate is based on all website visits by the household, rather than just category visits. For example, 41% of household 237's website visits prior to the attack are to Yahoo. This household visited two websites during the attack. Therefore the estimated probability of being in the treatment group is 0.41+0.41(1-0.41) =0.65.

The last measure repeats the regressions used in the main definition but the treatment group in defined by the category rather than all website visits. These regressions were only based on visits to the category rather than on the entire data set.

The bottom part of Table 5b shows the correlation coefficients of the main treatment definition and each of the three other definitions. The visit propensity measure and the category measure are highly correlated with the main treatment group definition. This is not surprising since all three measures are based on the individual-level propensity to visit the attacked website. The measure based on visits to rival websites is not highly correlated with the main measure.

2. Total Effect of the Denial of Service Attacks

2.1 Model and Identification

The decision to visit the attacked website is modeled as a discrete choice problem. I assume that Internet users choose the website that will give them the highest value on any particular choice occasion. The value of visiting a website is then:

$$(1) \qquad u_{ijt}= \gamma T_{ij}+\delta D_{jt}+\boldsymbol{\lambda D_{jt} T_{ij}}+X_{ijt}\beta+\mu_{ij}+\varepsilon_{ijt}$$

Here T_{ij} is the probability of being in the treatment group, D_{jt} is a vector representing a spline of the number of days since the attack occurred,[7] X_{ijt} is a vector of the other covariates included in the model (media mentions, choice last time, bytes uploaded, and an overall time trend), μ_{ij} is the household-level brand preference, and ε_{ijt} is a normally distributed idiosyncratic error term.

This framework explicitly allows the treatment group to have different preferences than the control group. It also allows for preferences to change over time. The treatment effect is therefore identified by the coefficient vector λ. This vector will capture the main effect as well as the decay of the effect over time. This is a difference-in-difference identification strategy.[8] The other covariates function as controls that allow for identification of the (treatment) effect of the website shutdown caused by the denial of service attacks. Since the data set contains only three months of data, long-run switching costs are subsumed into the household-level effect μ_{ij}.

I estimate this function using a random effects probit model. The dependent variable is a dummy variable for whether the household visits the website hit with the DoS attack (Yahoo, CNN, or Amazon) on a particular visit. For example, when estimating the impact of a DoS attack on Yahoo users, $y_{it}=1$ when a household visits Yahoo, and $y_{it}=0$ otherwise. This variable will equal one if $u_{ijt}\geq 0$ and it will equal zero otherwise. The coefficient μ_{ij} is assumed to be distributed i.i.d. Normal with mean μ_j and variance σ_j. The vector X_{ijt} includes the number of the previous 15 days in which the website was mentioned in the media, the previous experience at the attacked website in terms of the log of bytes uploaded to the website on the previous visit, an overall time trend, and whether the household chose the attacked website on the previous choice occasion.[9] Media mentions and previous experience are included as other factors that may affect choice. Media mentions is expected to have a positive impact on visit propensity. The website visited on the previous choice occasion is included to avoid mixing the overall loyalty effect with the direct effect of the DoS attack. Typically, this variable has a strongly positive effect on visit behavior. Without this variable, the measured effect of the attacks increases. Bytes uploaded will likely have a negative effect on Yahoo visits because it implies more effort was spent on the previous search. It will likely have a positive effect on CNN and Amazon visits because more time spent suggests more was achieved at the website. The time trend allows for changes to occur with time in the data set independent of the attacks. The econometric analysis identifies whether a household that experienced the attack changes its behavior relative to one that did not experience the attack.

I use a binary probit model rather than a multinomial logit (or multinomial probit) model because the binary model identifies the core coefficient with fewer assumptions. In particular, a probit specification does not require assumptions about the composition of the household-level choice set aside from the attacked website. The other brands in the choice set can change over time and across individuals.

This is particularly important for measuring the effect of unavailability on CNN and Amazon. While the top six websites in the portal category have 75% of the market, the top six news websites have only 51% and the top six shopping websites have only 28% (Table 3). Identifying a small number of websites for standard multinomial discrete choice analysis is not feasible. While there are sophisticated techniques for constructing choice sets (e.g. Andrews and Srinivasan, 1995), a binary model is much simpler and identifies the same phenomenon. All that matters for identification is whether the attacks changed visiting behavior at the attacked website. One weakness of the binary model is that changes in the choice set over time could lead to trends in both the mean and the variance of the alternative choice probabilities. In section 5, I show that results for Yahoo are robust to a multinomial logit specification.

2.2 Results: The Total Effect of the Denial of Service Attacks on the Attacked Sites

This section estimates the overall effect of the DoS attacks on Yahoo, CNN, and Amazon. Figure 1 shows the overall trends for those users who visited a rival website during the attacks. It plots market shares by week for the attacked websites, the rival websites visited during the attack instead of the attacked websites (calculated at the household level), and all other websites in the category for members of the treatment group. The numbers for the week of the attack are not included in the figure as they are lower as a direct consequence of the attacks.

Figure 1a shows that Yahoo's market share clearly dropped in the week following the denial of service attack. The market share of rival websites visited during the attack rises the following week. The market share of all other portals increases slightly. These effects appear to dissipate over time. Figures 1b and 1c show similar patterns for CNN and Amazon. In all three cases, the overall trends suggest that the attacks hurt the attacked website in the weeks that followed and that the attacks helped competitors.

Table 6 shows the results of the random effects probit regressions described in section 2.1. The first seven rows present the coefficients on the overall effect of the attacks. The attacks significantly decrease visits to Yahoo, CNN, and Amazon. For Yahoo, the effect of the attack decreases over time.[10] For CNN and Amazon, the trend is less pronounced but the overall effect appears largest in the days immediately following the attacks. The remaining rows present controls. The total effect of the attacks in lost visits is presented at the bottom of the table.

While the effect of unavailability due to the DoS attacks dissipates over time, it still has important economic implications. The results suggest that Yahoo lost an estimated 7.56 million visits between the attacks and the end of the sample (53 days later).[11] As mentioned earlier, Yahoo would have to make this up to advertisers. At four cents per visit, this totals $302,277. The medium-term effect of Yahoo being unavailable is over 3.4 times the short-term impact of $88,854 shown in Table 2.

In summary, the inability to access Yahoo, CNN, and Amazon had a lasting impact. In the case of Yahoo, this impact decreased over time and cost 3.4 times the immediate impact.

2.3 Results: The Total Effect of the Denial of Service Attacks on the Rival Sites

This section estimates the total impact of the DoS attacks on the rival websites that were visited during the attack instead of the attacked website. This effect is the sum of the benefit due to changing preferences and the benefit due to lock-in. Using the estimation strategy described in section 2.1, Table 7 shows that rivals visited during the attacks at Yahoo and Amazon gained from the attacks. The estimated effect of the attacks on CNN, though large, is generally not significantly different from zero.

The rival website visited instead of Yahoo is estimated to have gained 5.42 million visits or 72% of the visits that Yahoo lost. Not only did website unavailability hurt the attacked website, but the rival visited during the attack appears to have gained disproportionately. Section 3.1 describes a method for formally separating the roles of changing preferences and lock-in. Section 3.2 shows that lock-in mattered a great deal in the aftermath of the Yahoo attack. It mattered less in the Amazon attack. There is no statistically significant lock-in from the CNN attack.

3. Lock-in or Changing Preferences?

3.1 Model and Identification

In this section, I explain the strategy for identifying lock-in (switching costs). In section 3.2, I show the results. Lock-in to the rival website visited during the attack is identified by the difference in visit propensity between the rival website visited during the attacks and all other competitors. The lock-in resulting from the attacks will only affect rival websites that the users visit during the attack. All other websites competing with the attacked website can only benefit from a change in preferences (or a loss of lock-in at the attacked website).

The lock-in identified is then that which accrues at the rival website visited instead of the attacked website during the attack. For example, suppose household *i* visits MSN instead of Yahoo during the attack on Yahoo. MSN benefits from lock-in if it gains proportionately more than Altavista and Lycos as a consequence of the attack. Otherwise, the gain to MSN is just a consequence of the loss to Yahoo.

The value from visiting a website is defined as in equation (1),

$$u_{ijt}= T_{ij}\gamma+\delta D_{jt}+\boldsymbol{\lambda D_{jt}T_{ij}}+X_{ijt}\beta+\mu_{ij}+\varepsilon_{ijt}$$

The key to the identification of lock-in at the websites visited during the attacks is that the vector $\lambda D_{ij}A_{jt}$ will have a different meaning for rival websites that were visited during the attack and those that were not. The value of returning to the rival website that was visited during the attack will have a lock-in component. Other competing websites to the attacked website will not benefit from lock-in. They will only benefit from the reduced propensity to visit the attacked website. Therefore the value from visiting the attacked website for a household that experienced the attack is:

$$(2)\qquad u_{iat}= \gamma_{\boldsymbol{a}}T_{ia}+\delta_{\boldsymbol{a}}D_{\boldsymbol{at}}+\boldsymbol{\lambda_{a}D_{at}T_{ia}}+X_{iat}\beta+\mu_{ia}+\varepsilon_{iat}$$

Here the vector λ_a is the preference change resulting from the attack combined with any decrease in lock-in associated with the attacked website. As mentioned earlier, long-run switching costs are subsumed into the household-level effect μ_{ia}. The value from visiting a rival website that was visited during the attack is:

$$(3)\qquad u_{irt}= \gamma_{r}T_{ir}+\delta_{r}D_{rt}+\boldsymbol{\lambda_{r}D_{rt}T_{ir}}+X_{irt}\beta+\mu_{ir}+\varepsilon_{irt}$$

Here λ_r is the added lock-in associated with having visited the website an extra time in the past due to the DoS attack. Finally, the value from visiting a competing website that was not visited during the attack is

$$u_{iot}= X_{iot}\beta+\mu_{io}+\varepsilon_{iot} \tag{4}$$

The DoS attack will not directly enter the value gained at a website that was neither attacked nor visited during the attack. The attack will only affect the probability of visiting these other websites through the impact on the attacked websites and the rival websites that were visited during the attack. Consequently, controlling for user behavior before the attacks, exploring whether users are more likely to visit rival websites visited during the attack than other competing websites identifies the coefficient vector on lock-in, $\boldsymbol{\lambda_r}$.

In particular, a household visits the website that was visited during the attack instead of another competing website if $u_{irt}\geq u_{iot}$. Rearranging terms, this means that the website visited during the attack is visited again if

$$\gamma_r T_{ir}+\delta_r D_{rt}+\boldsymbol{\lambda_r D_{rt} T_{ir}}+(X_{irt} -X_{iot})\beta+\mu_{ir}-\mu_{io}+\varepsilon_{irt}-\varepsilon_{iot}\geq 0 \tag{5}$$

Therefore, estimating a probit model to see whether competing firms that were visited during the attacks gained more than other competing firms will identify the effect of lock-in, $\boldsymbol{\lambda_r}$.

This will allow for identification of short-run lock-in accruing to the rival website visited during the DoS attack. This method does not identify whether there is lock-in at the attacked websites. The identification assumes that households, on average, have accurate information about the quality of websites. For example, instead of lock-in, it could be that users systematically underestimate the value derived from the website visited instead of the attacked website. Upon visiting the website, their image of the site improves and they become more likely to visit in the future.

I find some evidence that the image does not improve: the effect of the attacks is short-lived. While this suggests that the switching cost dissipates over time, it also suggests that the underlying value gained from the rival website did not change as a consequence of the visit. In other words, decay means there is reversion to the state that occurred before the attacks. User behavior did not change in the long run. Therefore the users' preferences before the attacks were on average correct, providing support for the assumption and the identification argument.

The model also does not include an outside good. Therefore, a reduction in preference for the attacked website does not mean a reduction in total visits. The competing websites receive the complete benefit. If total visits are reduced, then the lock-in identification holds but the overall effect will be underestimated.

It is important to remember that this method identifies a particular kind of lock-in: the impact of an exogenous one-time switch on the website that benefited from the switch. All that is required for the lock-in effect to exist is that the act of visiting a website once will increase the probability of visiting that website in the future, all else being equal. Having visited a website at some point in the past must therefore have a medium-term impact on the value gained from visiting that website in the future.[12] Nevertheless, such a finding of state dependence does not preclude a change in preferences as well.

An alternative method of measuring online lock-in is to use a "power law" (Johnson, Bellman, and Lohse, 2003). While the power law idea is a very useful concept, it does not work in the context of identifying lock-in through this natural experiment. The power law uses many visits to identify how lock-in accrues. The impact of a one-time switch due to a DoS attack cannot be explicitly modeled as part of a power law because the switch is observed once. It is possible to implicitly think of the measured lock-in due to a one-time switch as part of an underlying power law that drives all choices.

3.2 Results: Lock-in

Section 2.2 showed that users who found the attacked website unavailable were less likely to return. Section 2.3 showed that rival websites visited during the attacks on Yahoo and Amazon gained. This section identifies the importance of lock-in at the rival websites in this effect.

Table 8 presents the lock-in results using the method described in section 3.1. Columns (1) through (3) show the results under the main treatment group definition for Yahoo, CNN, and Amazon respectively. Rivals to Yahoo and Amazon appear to have benefited from lock-in accrued during the attacks. The effect on CNN rivals is also positive though not significant. In all three cases, the lock-in decays over time.

The lock-in for Yahoo is highest in the first two days following the attack. It then dissipates. After 11 to 15 days, the attacks do not have a significant effect on Yahoo visits. In contrast, Table 6 showed that the overall effect on Yahoo had not fully dissipated between 31 and 53 days (the end of the sample). The lock-in effect of being unable to visit a website is short-lived relative to the overall effect on the attacked website.

Still, Yahoo's rivals did benefit from lock-in, although they also gained from a change in preferences. Yahoo rivals gained an estimated 2.78 million visits from lock-in ($111,317 at four cents per visit). For comparison, Table 7 shows that these websites gained 5.42 million visits overall ($216,770). Therefore 51% of the gains to Yahoo's rivals came from lock-in.

Figure 2 illustrates the relative impact of lock-in and the change in preferences over time. In particular, it shows the effect of the attacks on Yahoo of rival visits per day. On the first day, most of the effect is due to lock-in. From 2 to 5 days, both lock-in and the change in preferences play an important role. From 6 to 10 days, the effect of lock-in starts to diminish. By 11 to 15 days after the attacks, lock-in has almost no effect on visits to Yahoo's rivals.

Amazon rivals, however, gained much less from lock-in. Contrasting the last lines of Tables 7 and 8 shows that only 13% (122,609 of 909,701 visits) of the gains to Amazon's rivals came from lock-in. Unlike the case in the Yahoo attacks, after Amazon was unavailable rivals gained mostly from a lower preference for Amazon.

Columns (4) through (6) use different treatment group definitions to show robustness of the Yahoo results. Column (4) defines the treatment group by the degree to which the visit that occurred during the DoS attack was out of character for the household. This is the "relative to rival frequency" definition described in section 1.2. The overall results do not change. Columns (5) and (6) show that the Yahoo results are robust to defining the treatment group as visit propensity and by being in the category at the time.

In summary, lock-in that accrued during the attacks did help Yahoo rivals. The effect of the attacks on Amazon rivals was mostly a result of changing preferences. In the immediate aftermath of the attacks, lock-in substantially increased revenues for Yahoo rivals. This lock-in, however, was short-lived. While the overall impact of the Yahoo attack lasted at least 53 days, the lock-in disappeared after 11 to 15 days.

The next section shows that which segments were most affected.

4. Lock-in by User Segment

This section explores whether responses to Yahoo's unavailability differ by type of customer. This is important for designing policy responses to unavailability. First, if some segments are not affected by the unavailability, they do not need to be targeted by any response. Second, if some segments are particularly likely to be affected by lock-in, then, to the extent possible, these individuals should be targeted with short-run promotional campaigns aimed at bringing them back. Third, if some segments are particularly likely to reduce their preferences for the attacked website, then marketing initiatives aimed at these individuals should focus on quickly compensating them and should emphasize steps taken to prevent recurrences.

Segmentation is difficult in this case because the data do not contain reliable demographic information. To preserve individual anonymity, the Plurimus data only contain demographic information at the census block level. Therefore, instead of demographic data, I use observed behavior over the course of the sample, splitting the data in three ways. First, I compare frequent and infrequent users of the Internet. These groups are separated by the median household's time online in the sample. Frequent users are likely a good approximation for experienced users (Nie and Erbring, 2000). Second, I split the sample (at the median household) by proportion of time online spent on email and chat. Third, I compare households that use Yahoo Mail with those that do not.

Tables 6, 7, and 8 were re-estimated for each segment. Table 9 shows the coefficients by segment. In Tables 6, 7, and 8, these were the coefficients contained in the first seven rows. There are interesting differences between segments.

Frequent and infrequent users who found Yahoo unavailable during the attacks decreased their likelihood of visiting Yahoo. However, the reasons for this decrease differ by segment. Infrequent users appear to be especially likely to have developed lock-in. Furthermore, this lock-in is still significant from 16 to 30 days after the attacks. The lock-in that accrued to frequent users is relatively small and dissipates quickly.[13]

Columns 3 through 6 suggest that email campaigns in response to DoS attacks may be a mistake. In particular, columns 3 and 4 of Table 9 show that users who spent a relatively low proportion of their time online doing email and chat were especially affected by Yahoo's unavailability. They were also much more likely to develop lock-in at rival websites. Column 5 shows that users with Yahoo email accounts were barely affected by the attacks. Users without Yahoo email accounts, however, were particularly likely to have a medium-term response to unavailability. This suggests that, as a response to unavailability, websites should focus on relatively infrequent users of email and chat to the extent possible. Therefore, the relatively inexpensive response of email apologies and explanations to Yahoo's customers alone is unlikely to be enough. The results suggest that promotions need to be visible on other websites or offline.

In summary, the attacks had different impacts on different segments. These differences have important consequences for how websites should respond to unavailability.

5. Multinomial Logit Model: A Robustness Check

The random effects probit model used in this paper is not the conventional method for analyzing choice in marketing. More typically, a multinomial logit model is used. As described earlier, I use the probit model because it makes fewer assumptions about the composition of the individual-level choice set. Furthermore, it has a much lower computational burden. In this section, I show that the results for the

Yahoo attack hold in a multinomial logit model estimated on 300 randomly selected households. I do not estimate a multinomial logit model for the CNN and Amazon attacks because of the difficulty in defining a consistent choice set across individuals in news and online shopping.

In the multinomial logit model, the value for visiting the attacked website (Yahoo) is defined as in equation (2):

$$u_{iat}= \gamma_{\boldsymbol{a}}T_{ia}+\delta_{\boldsymbol{a}}D_{\boldsymbol{at}}+\boldsymbol{\lambda_{a}D_{at}T_{ia}}+X_{iat}\beta+\mu_{ia}+\varepsilon_{iat}$$

The value for visiting another website, j, is defined as:

$$u_{ijt}= R_{ij}\times(\gamma_jT_{ij}+\delta_jD_{jt}+\boldsymbol{\lambda_jD_{jt}T_{ij}}+\phi+\eta_i)+X_{ijt}\beta+\mu_{ij}+\varepsilon_{ijt} \tag{6}$$

Here, R_{ij} is equal to one if website j is the rival website visited by user i during the DoS attack and zero otherwise; ϕ is the coefficient for a rival brand dummy. This dummy is included to control for the fact that the website visited during the attack may be preferred to other brands. This preference is assumed to be distributed normally across households (captured by η_i). Unlike the probit models, therefore, the effect on Yahoo and the effect on rivals can be simultaneously estimated. For computational reasons, the choice set was restricted to the top six portals: Yahoo, MSN, Netscape, Excite, AOL, and Altavista. If a user visited MSN, Netscape, Excite, AOL, or Altavista during the attacks, then that website is considered the rival for that user. Otherwise, the user is assumed not to have visited a rival during the attacks. Preferences for each website j, μ_{ij}, are assumed to be distributed normally across households. The model is therefore estimated by maximum likelihood as a standard mixed logit (Train, 2003).

The vector X_{ijt} now contains media mentions over the previous 15 days, whether the website was chosen last time, whether the user has an email account at the portal, the length of the previous visit to the website, whether the previous search was repeated, a "missing data" dummy for whether the individual had previously visited the portal in the sample, and website-specific time trends.[14] Goldfarb (2006) shows length of the previous visit and whether the previous search was repeated are effective predictors of portal choice in the Plurimus data and provides detailed descriptions of their derivation.

Table 10 shows the results of the multinomial logit specification. The first seven rows on the left side show that Yahoo was hurt by the attacks. Members of the treatment group were less likely to visit the website in the aftermath of the attack, and this effect decreases over time.

The first seven rows on the right side of the table show that the rival website visited during the attack benefited more than other competing brands. Therefore, Yahoo's unavailability generated lock-in at the rival websites visited during the attack. This lock-in, however, is no longer significant after 16 to 30 days.

The other coefficients in Table 10 serve as controls. As expected, the website chosen the previous time, the rival dummy, and whether the user has an email account are positively correlated with choice. Also as expected, last search repeated suggests a bad experience at the website and decreases the likelihood of returning. Media mentions and length of the last visit to the portal had no significant effect.

The multinomial logit specification shows similar results to the probit specification emphasized throughout the paper.

6. Discussion and Conclusions

This paper has measured the immediate and medium-term impact of one particular source of service unavailability: Internet denial of service attacks. It has shown the unavailability had both an immediate and a medium-term effect on Yahoo, CNN, and Amazon. In the case of Yahoo, the medium-term effect was 3.4 times larger than the immediate effect. While the effect decreased over time, it was still significant between 31 and 53 days. In total, the DoS attack on Yahoo cost it an estimated $391,131. This suggests that it would be worthwhile for Yahoo to spend a significant amount of money to prevent future shutdowns but no more than the above amount per expected attack. Given that the main method of attack prevention is to have programmers available to respond in real time (Gordon et al., 2004), this result suggests that Yahoo should maintain a small staff capable of responding to attacks when they occur.

This study also identified the lock-in resulting from users visiting a different website when the attacked website was not available. It found that lock-in drove 51% of the gain to rival websites visited instead of Yahoo during the attack. Lock-in only drove 13% of the gain to rivals from the Amazon attack. While more research is needed to see if this result holds in other contexts, this may suggest that short-run lock-in matters more at free websites such as portals than at ecommerce websites. Relative to the overall impact, however, the impact of lock-in was short-lived. While in the first days after the attack, lock-in accounted for much of the impact, the effect of lock-in dissipated within 11 to 15 days.

Different users had different reactions to unavailability. While Yahoo's unavailability had a similar overall effect on frequent and infrequent users, infrequent users developed relatively high lock-in to the rival website visited during the attacks. Users who visited relatively few email and chat websites were particularly affected if they could not access Yahoo during the attacks.

These results have important managerial implications. Since lock-in dissipates quickly, the main response to website unavailability should focus on the overall perception of the brand. The unavailability has an effect on the perceived value gained from visiting the website. Short run promotional campaigns aimed at overcoming lock-in will not be effective on most users.

The results also have implications on the promotional medium used in response to website unavailability. While responding to DoS attacks is strategically challenging due to their unpredictability, the relatively easy-to-implement option of emailing apologies and compensation is likely to be relatively ineffective. Yahoo email users and heavy users of email and chat websites were not strongly affected by the attacks. This provides some evidence that simply emailing Yahoo's email users may be a waste of resources. Promotional responses to unavailability may be most effective if visible on other websites or offline.

Overall, rival brands visited during the attacks gain little long run, or even medium run, advantage. While the attacks generate significant short run visits, lock-in dissipates after two weeks. Infrequent users, however, provide an important opportunity for rival brands (and a key threat to the attacked brands). These users are particularly likely to develop lock-in at a rival website as a consequence of unavailability. From a manager's perspective, this suggests that easily learned website features may be particularly helpful in leveraging a competitor's availability problems.

This study also makes a small contribution to the literature on whether consumers hold firms responsible for the damaging actions of other parties. Consumers reduce their preference for a website after a DoS attack. This result is similar to Mitchell's (1989) results on the Tylenol poisonings of 1982. He uses an event study to show that the incident led to a substantial decrease in the brand value of Tylenol.

Overall, this study examines just one instance of unavailability. It is important be cautious in generalizing the results, especially to offline shopping environments. There are differences between the online and offline contexts. First, in the case of a retail stockout, it is not clear how consumers will apportion blame to the retailer and the manufacturer. In the case of website unavailability, there is only one company. Second, websites are experience goods. The characteristics of consumer package goods, for example, may be more easily observed. Finally, there is considerable evidence that online switching costs, while non-trivial, are lower than offline switching costs (e.g. Goldfarb, 2006; Gandal, 2001). The magnitude of offline lock-in due to unavailability may be different.

In conclusion, this paper has developed a method to identify the relative importance of lock-in and changing preferences in the impact of product (or service) unavailability on future choices. The results suggest that the medium-term impact of unavailability is due to both a change in preference and lock-in. The identification method could easily be applied to stockouts in grocery stores. Comparing the impact of the stockout on the brand that is bought instead with other brands that are not bought will allow identification of lock-in in these markets. Future work should explore the consequences of unavailability on future choices in an offline setting.

References

Andrews, Rick L., and T.C. Srinivasan. (1995). "Studying Consideration Effects in Empirical Choice Models Using Scanner Panel Data", Journal of Marketing Research 32, 30-41.

Bell, David R., and Gavan J. Fitzsimons. (1999). "An Experimental and Empirical Analysis of Consumer Response to Stockouts", Working Paper. Wharton School, University of Pennsylvania.

Campo, Katia, Els Gijsbrechts, and Patricia Nisol. (2000). "Towards Understanding Consumer Response to Stock-Outs", Journal of Retailing 76, 219-242.

Corsten, Daniel, and Thomas Gruen. (2004). "Stock-Outs Cause Walkouts", Harvard Business Review 82, 26-28.

Coughlan, Anne T., Erin Anderson, Louis Stern, and Adel I. El-Ansary. (2001). Marketing Channels, 6th edition. Upper Saddle River, NJ: Prentice Hall.

Fitzsimons, Gavan J. (2000). "Consumer Response to Stockouts", Journal of Consumer Research 27, 249-266.

Gandal, Neil. (2001). "The dynamics of competition in the internet search engine market", International Journal of Industrial Organization 19, 1101-1117.

Goldfarb, Avi. (2006). "State Dependence at Internet Portals", forthcoming Journal of Economics and Management Strategy.

Gordon, Lawrence A., Martin P. Loeb, William Lucyshyn, and Robert Richardson. (2004). Ninth Annual CSI/FBI Computer Crime and Security Survey. San Francisco: Computer Security Institute.

Jeuland, Abel P. (1979). "The Interaction Effect of Preference and Availability on Brand Switching and Market Share", Management Science 25, 953-965.

Johnson, Eric J., Bellman, Steven, and Lohse, Gerald L. (2003). "Cognitive Lock-In and the Power Law of Practice", Journal of Marketing 57, 62-75.

Keane, Michael P. (1997). "Modeling Heterogeneity and State Dependence in Consumer Choice Behavior", Journal of Business and Economic Statistics 15, 310-327.

Manchanda, Puneet, Jean-Pierre Dube, Pradeep K. Chintagunta, and Khim Yong Goh. (2006). "The Effect of Banner Advertising on Internet Purchasing." Journal of Marketing Research 43, forthcoming.

Mela, Carl F., Sunil Gupta, and Donald R. Lehmann. (1997). "The Long-Term Impact of Promotion and Advertising on Consumer Brand Choice." Journal of Marketing Research 34, 248-261.

Nie, Norman H., and Lutz Erbring. 2000. "Internet and Society: A Preliminary Report." Stanford Institute for the Quantitative Study of Society. Mimeographed.

Sandoval, Greg, and Troy Wolverton. (2000). "Leading Web sites under attack." CNET News.com. February 9, posted 1:50PM PT. http://news.com.com/2100-1017-236683.html.

Seetharaman, P. B., Andrew Ainslie, and Pradeep Chintagunta. (1999). "Investigating Household State Dependence Effects Across Categories", Journal of Marketing Research 36, 488-500.

Swait, Joffre, and Tulin Erdem. (2002). "The Effects of Temporal Consistency of Sales Promotions and Availability on Consumer Choice Behavior", Journal of Marketing Research 39, 304-320.

Train, Kenneth. (2003). Discrete Choice Methods with Simulation. Cambridge UK: Cambridge University Press.

Yahoo Sales. (2000). Personal correspondence. April 12.

Yahoo, Inc. (2005). "Terms of Service", http://docs.yahoo.com/info/terms/. Viewed on May 3, 2005.

Table 1: Comparison of Plurimus Data with Nielsen/NetRatings Data (February 2000)

	Plurimus Data	Nielsen/NetRatings Home Users[a]
Top 10 Web Properties[b]		
1	Yahoo	AOL
2	MSN[c]	Yahoo
3	AOL	MSN[c]
4	Lycos	Excite
5	Excite	Lycos
6	Go	Go
7	Altavista	Time Warner
8	Amazon	NBCi
9	NBCi	Amazon
10	Time Warner	Altavista
Average time online per month	9 hours 58 minutes	9 hours 19 minutes

[a] source: Nielsen/NetRatings reproduced March 13, 2005 at http://cyberatlas.internet.com/big_picture/demographics/article/0,1323,5931_322381,00.html

[b] Web properties include all websites run by the same organization. For example, YahooNews and YahooSports are part of the Yahoo property and Disney and ESPN are part of the Go property.

[c] MSN includes both Microsoft.com and MSN.com.

Table 2: Timing of the Attacks

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Website	Time of Attack[a]	# users in category at time	# users online at time	# users online at time who visited the site before the attacks	Estimated # users in the treatment group	Estimated Immediate Lost Unique Visits[b]	Estimated Immediate Revenue Impact[b]
Yahoo	Mon. Feb. 7: 1:20 PM–4:20 PM	401	650	525	136	2,221,350	$88,854
CNN	Tues. Feb. 8: 7:00 PM–8:50 PM	56	587	229	40	653,338	$26,134
Amazon	Tues. Feb. 8: 8:00 PM–9:00 PM	38	423	210	32	522,671	$20,907
EBay	Tues. Feb. 8: 6:20 PM–7:50 PM	10	375	151	20	326,669	$13,067
ZDNet	Wed. Feb. 9: 6:45 AM–9:45 AM	16	397	181	11	179,668	$7,187
Buy.com	Tues. Feb. 8: 1:50 PM–4:50 PM	88	717	52	2	32,667	$1,307
E*Trade	Wed. Feb. 9: 8:00 AM–9:30 AM	37	168	29	1.2	19,600	$784

[a] All times EST. Source: CNET (Sandoval and Wolverton 2000).

[b] The market size of 43.3 million online households is based on Plurimus Corporation estimates. The revenue estimates assume 4 cents per visit. This value is based on revenue estimates from J. Walter Thompson Company for nine portals from January to March of 2000 combined with the visits data in this study. The revenue estimate is based on data for Internet portals, and will be most reliable for Yahoo.

Table 3: Competitive Sets and Market Shares

	Yahoo	CNN	Amazon
	(Portals)	(News)	(Shopping)
Number of Websites	140	182	366
Number of Observations	855,370	106,129	69,342
Number of Users	2,479	1,544	1,932
Own Share	33.00%	8.22%	14.02%
Top Ten Competitors and their shares			
1	*MSN*	*MSNBC*	*YahooShopping*
	17.42%	15.44%	3.34%
2	*Netscape*	*USA Today*	*CD Now*
	10.60%	9.12%	3.05%
3	*Excite*	*YahooNews*	*eNews*
	5.26%	8.51%	2.98%
4	*AOL*	*WRAL*	*BN.com*
	4.29%	4.81%	2.85%
5	*Altavista*	*News & Observer*	*Shopping.com*
	3.94%	4.68%	1.69%
6	*iWon*	*Nando.net*	*Buy.com*
	2.62%	2.93%	1.56%
7	*Lycos*	*Post-Gazette*	*Shopnow*
	2.55%	2.91%	1.46%
8	*Myway*	*NY Times*	*MSN eShops*
	2.13%	2.49%	1.44%
9	*Go*	*ABC News*	*Spree.com*
	1.98%	2.35%	1.21%
10	*Hotbot*	*Drudge Report*	*Columbia House*
	1.85%	2.23%	0.96%

Table 4: Descriptive Statistics

	Yahoo	CNN	Amazon
# Media Mentions over past 15 days			
Mean	6.996	1.479	6.501
Standard Deviation	2.636	1.652	2.097
Minimum	0	0	0
Maximum	13	6	11
Log(bytes uploaded on last visit to attacked site)			
Mean	7.573	8.7413	8.783
Standard Deviation	1.082	1.264	1.371
Minimum	0	4.595	0
Maximum	15.072	12.774	13.313
Probability in treatment group (main definition)			
Mean	0.157	0.125	0.092
Standard Deviation	0.279	0.085	0.031
Minimum	0	0	0
Maximum	0.998	0.885	0.946
Probability in treatment group (relative to rival frequency)			
Mean	0.773	0.810	0.939
Standard Deviation	0.366	0.176	0.098
Minimum	0	0	0.023
Maximum	1	1	1
Probability in treatment group (relative to visit propensity)			
Mean	0.201	0.132	0.085
Standard Deviation	0.338	0.176	0.054
Minimum	0	0	0
Maximum	0.999	0.934	0.878
Probability in treatment group (in category at the time)			
Mean	0.172	0.102	0.056
Standard Deviation	0.322	0.0728	0.025
Minimum	0	0	0
Maximum	1	0.810	0.629
# Observations	855,370	106,129	69,342

Table 5a: Identifying Treatment Groups: Fixed Effect Probit Results

	Yahoo	CNN	Amazon
Choose last time	0.360 (8.75E-03)**	0.128 (0.041)**	0.315 (0.055)**
# Media Mentions over past 15 days	0.052 (0.024)*	0.0123 (0.0972)	0.0329 (0.0168)+
Log(bytes uploaded on last visit to attacked site)	0.056 (2.22E-03)**	0.089 (6.40E-03)**	0.124 (5.51E-03)**
Day[a]	0.0191 (0.0214)	-0.00114 (0.00615)	0.0981 (0.0865)
# Observations	1,124,894	350,551	782,233
LL	-216,995	-14,470	-16,320

All regressions include household fixed effects and hour of the day dummies.
Standard Errors in parentheses.
[a]coefficients multiplied by 10,000
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 5b: Identifying Treatment Groups: Descriptives

		Yahoo	CNN	Amazon
A) Distribution of Probability in treatment group[a]				
	10th percentile	0.017	0.010	0.007
	25th percentile	0.045	0.044	0.027
	50th percentile	0.149	0.097	0.077
	75th percentile	0.393	0.267	0.179
	90th percentile	0.741	0.582	0.427
B) Comparison of Pr(treatment)>0 and Pr(treatment)=0				
% of households in category sample with positive probability of being in treatment group		20.2%	11.8%	9.2%
	Positive Probability in treatment group	13.81	14.24	13.96
Average years education in census block		(0.062)	(0.125)	(0.096)
(standard errors in parentheses)	Not in treatment group	13.97	13.96	13.93
		(0.032)	(0.032)	(0.031)
	Positive Probability in treatment group	$46,671	$49,806	$46,907
Average income in census block		($1,015)	($1,956)	($1,343)
(standard errors in parentheses)	Not in treatment group	$48,103	$47,527	$47,735
		($492)	($491)	($486)
	Positive Probability in treatment group	10.01%	9.30%	10.85%
% time online spent in email or chat		(0.449%)	(0.772%)	(0.708%)
(standard errors in parentheses)	Not in treatment group	7.55%	8.03%	7.75%
		(0.212%)	(0.214%)	(0.203%)
	Positive Probability in treatment group	89.1	115.4	108.6
Hours online (average by household)		(4.83)	(10.8)	(9.38)
(standard errors in parentheses)	Not in treatment group	45.6	69.7	63.5
		(0.898)	(1.42)	(1.14)
C) Actual # users same time & day of week as attack				
	week 1	140	33	27
	week 2	133	37	23
	week 3	152	44	39
	week 4	144	41	31
	week 5	127	34	26
	Predicted # users at attacked site during attack	136	40	32
D) Correlation coefficient of various treatment definitions				
	Correlation with treatment relative to rival frequency[a]	0.057	0.012	-0.057
	Correlation with treatment by visit propensity[a]	0.911	0.898	0.744
	Correlation with treatment defined by category-level data[a]	0.651	0.724	0.662

[a] Conditional on strictly positive probability in treatment group.

Table 6: Impact of the Denial of Service Attacks on the Attacked Websites Random Coefficient Probit Regression Results

	(1)	(2)	(3)
	Yahoo	CNN	Amazon
1 Day after attack and in treatment group	-0.191	-5.097	-4.838
	(0.0590)**	(3.034)+	(1.317)**
2 Days after attack and in treatment group	-0.169	-1.230	-4.073
	(0.0752)*	(0.562)*	(2.191)+
3 to 5 Days after attack and in treatment group	-0.176	-0.248	-1.716
	(0.0518)**	(0.302)	(1.018)+
6 to 10 Days after attack and in treatment group	-0.154	-0.750	-5.764
	(0.0437)**	(0.240)**	(1.672)**
11 to 15 Days after attack and in treatment group	-0.143	-0.605	-0.654
	(0.0390)**	(0.234)**	(0.781)
16 to 30 Days after attack and in treatment group	-0.0340	0.130	-1.450
	(0.0256)	(0.186)	(0.737)*
Over 30 Days after attack and in treatment group	-0.00961	-0.207	-2.150
	(0.00230)**	(0.160)	(0.558)**
1 Day after attack	-0.0271	-1.305	-0.309
	(0.0309)	(0.407)**	(0.200)
2 Days after attack	-0.0124	0.00742	-0.134
	(0.0240)	(0.0919)	(0.0757)+
3 to 5 Days after attack	0.0164	-0.0269	-0.0502
	(0.0149)	(0.0621)	(0.0446)
6 to 10 Days after attack	0.0473	0.188	-0.0171
	(0.0116)**	(0.0448)**	(0.0409)
11 to 15 Days after attack	0.0494	0.0797	-0.172
	(0.0132)**	(0.0528)	(0.0398)**
16 to 30 Days after attack	0.0208	0.0300	-0.0967
	(0.00876)*	(0.0298)	(0.0232)**
Over 30 Days after attack	0.0251	0.0926	-0.0604
	(0.00702)**	(0.0278)**	(0.0214)**
Treatment group	0.0953	0.466	0.216
	(0.0201)**	(0.106)**	(0.0431)**
# Media Mentions over past 15 days	0.000215	-0.00133	0.00632
	(0.00129)	(0.00757)	(0.00544)
Choose last time	1.137	1.240	1.007
	(0.00481)**	(0.0216)**	(0.0177)**
Log(bytes uploaded on last visit to attacked site)	-0.0308	-0.0367	0.0219
	(0.00219)**	(0.00869)**	(0.00601)**
Day[a]	0.00514	0.0788	0.0691
	(0.0122)	(0.0370)*	(0.0420)
Constant (Mean)	-0.972	-1.891	-1.712
	(0.0268)**	(0.0973)**	(0.0952)**
Constant (Standard Deviation)	0.732	0.728	0.462
	(0.00566)**	(0.0180)**	(0.0611)**
Observations	855,370	106,136	69,342
# Users	2,601	1,946	2,287
LL	-261,127	-13,086	-21,021
Simulated lost visits[b]	7,556,917	307,547	1,310,058

Standard errors in parentheses.
+p<0.10; *p<0.05; **p<0.01
[a] coefficients multiplied by 10,000; [b] Based on 43.3 million online households

Table 7: Total Impact of the Attacks on Rival Websites
Random Coefficient Probit Regression Results Using the Main Treatment Definition

	(1) Yahoo	(2) CNN	(3) Amazon
1 Day after attack and in treatment group	0.385	0.728	0.974
	(0.139)**	(0.553)	(0.551)+
2 Days after attack and in treatment group	0.254	0.232	0.134
	(0.151)+	(0.188)	(0.578)
3 to 5 Days after attack and in treatment group	0.112	0.312	1.511
	(0.0685)	(0.137)*	(0.569)**
6 to 10 Days after attack and in treatment group	0.0933	0.210	0.907
	(0.0614)*	(0.103)*	(0.397)*
11 to 15 Days after attack and in treatment group	0.0221	0.0748	1.233
	(0.00618)**	(0.108)	(0.503)*
16 to 30 Days after attack and in treatment group	0.0163	-0.0360	0.347
	(0.0373)	(0.0628)	(0.334)
Over 30 Days after attack and in treatment group	-0.0390	-1.006	0.483
	(0.0344)	(0.748)	(0.318)
1 Day after attack	-0.0993	-0.349	0.636
	(0.0516)+	(0.260)	(0.615)
2 Days after attack	0.0850	0.132	0.468
	(0.0365)*	(0.185)	(0.247)+
3 to 5 Days after attack	0.108	0.275	0.0251
	(0.0244)**	(0.105)**	(0.194)
6 to 10 Days after attack	-0.0805	-0.405	0.266
	(0.0202)**	(0.0795)**	(0.133)*
11 to 15 Days after attack	-0.149	-0.193	0.116
	(0.0211)**	(0.0914)*	(0.163)
16 to 30 Days after attack	-0.0241	0.133	0.238
	(0.0133)+	(0.0584)*	(0.0955)*
Over 30 Days after attack	-0.0461	0.0765	-0.110
	(0.0120)**	(0.0498)	(0.0947)
Treatment Group	-0.179	-0.190	-0.596
	(0.0313)**	(0.191)	(0.390)
Chose rival last time	0.8233	0.901	0.656
	(0.00769)**	(0.0375)**	(0.0656)**
Log(bytes uploaded on last visit to rival site)	0.00676	0.0117	0.0112
	(0.00192)**	(0.0106)	(0.00868)
Day[a]	0.0203	-0.00623	-0.0270
	(0.0200)	(0.0704)	(0.162)
Constant (Mean)	-0.441	-0.0795	-1.235
	(0.0356)**	(0.168)	(0.266)**
Constant (Std. Deviation)	0.563	0.952	0.989
	(0.0107)**	(0.102)**	(0.127)**
Observations	309,413	22,504	5,719
# Users	401	56	38
LL	-109,677	-4,551	-1,351
Simulated Rival gain due to the attacks (visits)[b]	5,419,241	283,049	909,701

Standard errors in parentheses.
+p<0.10; *p<0.05; **p<0.01
[a]coefficients multiplied by 10,000; [b]Based on 43.3 million online households

Table 8: Impact of the Attacks on Rival Websites Relative to Website that were not Attacked Identifying Lock-In Using Random Coefficient Probit Regressions

	Main Treatment Definition			Treatment Relative to Rival Frequency	Treatment as visit propensity	Treatment in category at the time
	(1)	(2)	(3)	(4)	(5)	(6)
	Yahoo	CNN	Amazon	Yahoo	Yahoo	Yahoo
1 Day after attack and in treatment group	0.410	0.521	0.754	0.566	0.292	0.314
	(0.164)*	(0.993)	(0.371)*	(0.138)**	(0.116)*	(0.115)**
2 Days after attack and in treatment group	0.172	0.0346	0.232	0.274	0.1255	0.0722
	(0.0842)*	(0.293)	(0.589)	(0.108)*	(0.0845)	(0.0842)
3 to 5 Days after attack and in treatment group	0.0934	-0.213	0.375	0.157	0.0606	0.108
	(0.0498)+	(0.156)	(0.181)*	(0.0712)*	(0.0554)	(0.0551)*
6 to 10 Days after attack and in treatment group	0.0808	0.160	0.357	0.0551	0.153	0.205
	(0.0405)*	(0.115)	(0.193)+	(0.00598)**	(0.0489)**	(0.0486)**
11 to 15 Days after attack and in treatment group	0.0159	0.0705	0.110	0.0541	0.053	0.0521
	(0.0131)	(0.120)	(0.0529)*	(0.00607)**	(0.0497)	(0.0492)
16 to 30 Days after attack and in treatment group	-0.0205	0.0384	-0.137	0.0281	-0.0318	0.0195
	(0.0420)	(0.727)	(0.364)	(0.00385)**	(0.0301)	(0.0296)
Over 30 Days after attack and in treatment group	-0.0363	-0.0619	0.120	-0.0206	-0.0517	0.000172
	(0.0389)	(0.0622)	(0.354)	(0.0348)	(0.0365)	(0.0258)
1 Day after attack	-0.0567	-0.363	0.726	-0.148	-0.215	-0.247
	(0.0545)	(0.261)	(0.628)	(0.0624)*	(0.105)*	(0.105)*
2 Days after attack	0.0844	0.164	0.495	0.0420	0.0951	0.0588
	(0.0383)*	(0.193)	(0.252)*	(0.0429)	(0.0781)	(0.0779)
3 to 5 Days after attack	0.101	0.249	0.00228	0.0426	0.0287	-0.00816
	(0.0256)**	(0.105)*	(0.203)	(0.0276)	(0.0511)	(0.0509)
6 to 10 Days after attack	-0.0426	-0.421	0.281	-0.196	-0.170	-0.206
	(0.0213)*	(0.0807)**	(0.138)*	(0.0242)**	(0.0456)**	(0.0453)**
11 to 15 Days after attack	-0.110	-0.211	0.103	-0.251	-0.233	-0.266
	(0.0222)**	(0.0929)*	(0.172)	(0.0238)**	(0.0463)**	(0.0460)**
16 to 30 Days after attack	0.0119	0.125	0.305	-0.116	-0.00744	-0.0456
	(0.0139)	(0.0596)*	(0.101)**	(0.0153)**	(0.0281)	(0.0277)+
Over 30 Days after attack	-0.00618	0.106	-0.106	-0.143	-0.0401	-0.0787
	(0.0126)	(0.0513)*	(0.0100)	(0.0138)**	(0.0246)	(0.0241)**
Treatment group	-0.157	-0.0719	-0.354	-0.267	-0.259	-0.243
	(0.0410)**	(0.189)	(0.499)	(0.0326)**	(0.0272)**	(0.0203)**
Chose rival last time	0.815	0.872	0.583	0.813	0.815	0.835
	(0.00809)**	(0.0392)**	(0.0710)**	(0.00813)**	(0.00812)**	(0.00801)**
Log(bytes uploaded on last visit to rival site)	0.0115	0.0153	0.00944	0.00443	0.0103	0.0206
	(0.00217)**	(0.0109)	(0.00893)	(0.00229)+	(0.00216)**	(0.00199)**
Day[a]	-0.00102	0.0216	-0.0374	-0.00334	0.00242	0.00181
	(0.0208)	(0.0733)	(0.164)	(0.0215)	(0.0208)	(0.00227)
Constant (Mean)	-0.306	-0.157	-1.252	-0.102	-0.355	-0.312
	(0.0385)**	(0.172)	(0.287)**	(0.0391)**	(0.0397)**	(0.0374)**
Constant (Std. Deviation)	0.639	0.954	1.255	0.629	0.639	0.526
	(0.0130)**	(0.102)**	(0.192)**	(0.0145)**	(0.0135)**	(0.00942)**
Observations	221,842	21,828	5,530	221,842	221,842	221,842
# Users	401	56	38	401	401	401
LL	-89,746	-4,253	-1,300	-89,688	-89,795	-89,949
Simulated Rival Gain due to Lock-in (visits)[b]	2,782,923	82,746	122,609	4,642,118	3,418,916	4,709,400

Standard errors in parentheses. +p<0.10; *p<0.05; **p<0.01.

[a]coefficients multiplied by 10,000; [b]Based on 43.3 million online households

Table 9: Results by Segment after the Attacks (Yahoo only)

		Infrequent Users	Frequent Users	High Email and Chat	Low Email and Chat	Use Yahoo Mail	Don't Use Yahoo Mail
Overall Effect on Yahoo (as Table 6)	1 Day after attack and in treatment group	-0.208 (0.118)+	-0.175 (0.081)*	-0. 124 (0.0734)+	-0.218 (0.102)*	-0.127 (0.150)	-0.337 (0.145)*
	2 Days after attack and in treatment group	-0.172 (0.0535)**	-0.142 (0.0810)+	-0.140 (0.0758)+	-0.177 (0.0928)+	-0.0882 (0.119)	-0.193 (0.101)+
	3 to 5 Days after attack and in treatment group	-0.180 (0.0997)+	-0.150 (0.0634)*	-0.0972 (0.0653)	-0.169 (0.0887)+	-0.149 (0.0818)+	-0.167 (0.0681)*
	6 to 10 Days after attack and in treatment group	-0.201 (0.0866)*	-0.152 (0.0534)**	-0.0486 (0.0534)	-0.160 (0.0836)+	-0.0914 (0.0623)	-0.173 (0.0647)**
	11 to 15 Days after attack and in treatment group	-0.124 (0.0467)**	-0.203 (0.0470)**	-0.0806 (0.0480)+	-0.153 (0.0780)*	-0.0673 (0.0583)	-0.164 (0.0570)**
	16 to 30 Days after attack and in treatment group	-0.0269 (0.0517)	-0.0209 (0.0316)	-0.0471 (0.0316)	-0.0903 (0.0475)+	-0.0450 (0.0381)	-0.128 (0.0361)**
	Over 30 Days after attack and in treatment group	-0.0207 (0.0484)	-0.0516 (0.0285)+	-0.0260 (0.0692)*	-0.0459 (0.0917)*	-0.0497 (0.0349)	-0.0657 (0.0324)*
Overall Effect on Rival (as Table 7)	1 Day after attack and in treatment group	0.545 (0.140)**	0.217 (0.111)*	0.418 (0.169)*	0.669 (0.259)**	0.225 (0.135)+	0.406 (0.164)*
	2 Days after attack and in treatment group	0.393 (0.179)*	0.141 (0.0811)+	0.123 (0.119)	0.337 (0.164)*	0.136 (0.194)	0.279 (0.120)*
	3 to 5 Days after attack and in treatment group	0.149 (0.0648)*	0.0540 (0.0807)	0.0875 (0.0870)	0.139 (0.114)	0.151 (0.0561)**	0.179 (0.0816)*
	6 to 10 Days after attack and in treatment group	0.0954 (0.0314)**	0.102 (0.0697)	0.168 (0.0724)*	0.383 (0.124)**	0.111 (0.118)	0.137 (0.0777)+
	11 to 15 Days after attack and in treatment group	0.0256 (0.0137)+	0.0239 (0.00706)**	0.162 (0.0736)*	0.350 (0.120)**	0.0178 (0.0123)	0.0378 (0.0172)*
	16 to 30 Days after attack and in treatment group	0.0116 (0.0785)	0.0309 (0.0438)	0.00472 (0.0456)	0.0463 (0.0680)	0.0382 (0.0715)	0.0134 (0.0462)
	Over 30 Days after attack and in treatment group	-0.0104 (0.0344)	-0.0405 (0.0901)	-0.0979 (0.0425)*	-0.0285 (0.0610)	-0.0294 (0.0672)	-0.0146 (0.0428)
Relative Effect on Rival (as Table 8)	1 Day after attack and in treatment group	0.867 (0.317)**	0.203 (0.106)+	0.336 (0.196)+	0.722 (0.175)**	0.119 (0.0547)*	0.423 (0.203)*
	2 Days after attack and in treatment group	0.485 (0.139)**	0.132 (0.137)	0.190 (0.141)	0.205 (0.102)*	0.0942 (0.0522)+	0.100 (0.0312)**
	3 to 5 Days after attack and in treatment group	0.128 (0.518)*	0.0989 (0.0925)	0.0989 (0.0321)**	0.138 (0.0451)**	0.0436 (0.0182)*	0.186 (0.0979)+
	6 to 10 Days after attack and in treatment group	0.141 (0.0793)+	0.0822 (0.0184)**	0.0731 (0.0820)	0.146 (0.145)	0.0804 (0.130)	0.171 (0.0902)+
	11 to 15 Days after attack and in treatment group	0.0396 (0.0175)*	0.0127 (0.0777)	0.0134 (0.00827)	0.0405 (0.0133)**	0.0184 (0.0133)	0.269 (0.0873)**
	16 to 30 Days after attack and in treatment group	-0.0185 (0.0103)+	-0.0166 (0.0482)	-0.0211 (0.0512)	-0.0268 (0.0792)	-0.0286 (0.0804)	-0.0164 (0.0541)
	Over 30 Days after attack and in treatment group	-0.0289 (0.0980)	-0.0316 (0.0444)	-0.0319 (0.0476)	-0.0486 (0.0716)	-0.0728 (0.0765)	-0.0248 (0.0503)

Regressions are random effect probit regressions and include the same covariates as Tables 6, 7, and 8.
Coefficients shown. Standard errors in parentheses.
+p<0.10; *p<0.05; **p<0.01

Table 10: Multinomial Logit Results for Yahoo

		Coefficient	Standard Error			Coefficient	Standard Error
Interactions with Yahoo	1 Day after attack and in treatment group	-1.014	0.189**	Interactions with Rival Brand visited during attacks	1 Day after attack and in treatment group	0.328	0.0758**
	2 Days after attack and in treatment group	-0.855	0.409*		2 Days after attack and in treatment group	0.166	0.0606**
	3 to 5 Days after attack and in treatment group	-0.204	0.0652**		3 to 5 Days after attack and in treatment group	0.236	0.0406**
	6 to 10 Days after attack and in treatment group	-0.626	0.219**		6 to 10 Days after attack and in treatment group	0.114	0.0395**
	11 to 15 Days after attack and in treatment group	-0.0946	0.0202**		11 to 15 Days after attack and in treatment group	0.0733	0.0407+
	16 to 30 Days after attack and in treatment group	-0.0505	0.0136**		16 to 30 Days after attack and in treatment group	-0.0616	0.0725
	Over 30 Day after attack and in treatment group	-0.0667	0.0119**		Over 30 Day after attack and in treatment group	-0.0915	0.122
	1 Day after attack	0.133	0.129		1 Day after attack	-0.0439	0.0211*
	2 Days after attack	-0.0444	0.124		2 Days after attack	-0.0206	0.164
	3 to 5 Days after attack	-0.0251	0.0816		3 to 5 Days after attack	0.0625	0.115
	6 to 10 Days after attack	0.155	0.0635*		6 to 10 Days after attack	-0.107	0.0915
	11 to 15 Days after attack	0.166	0.0655*		11 to 15 Days after attack	-0.316	0.0946**
	16 to 30 Days after attack	0.190	0.0623**		16 to 30 Days after attack	0.134	0.0641*
	Over 30 Days after attack	0.288	0.0798**		Over 30 Days after attack	0.214	0.0577**
	Treatment group	0.248	0.0737**		Treatment group	-0.251	0.126*
					Rival Dummy (Mean)	1.561	0.0328**
Brand Dummies[a]	Yahoo (Mean)	0.839	0.0479**		Rival Dummy (Std. Dev.)	0.765	0.0899**
	Yahoo (Std. Dev.)	0.106	0.0253**				
	MSN (Mean)	0.675	0.0391**	Variables used for all brands	Chose website last time	1.692	0.0204**
	MSN (Std. Dev.)	0.324	0.149*		Person has email account on the portal	0.293	0.0120**
	Netscape (Mean)	0.517	0.0412**		Log(Length Last Visit)	-0.000908	0.00721
	Netscape (Std. Dev.)	0.277	0.110*		Last Search Repeated	-0.840	0.0176**
	Excite (Mean)	-0.315	0.0501**		Missing Data	-0.664	0.0373**
	Excite (Std. Dev.)	0.0453	0.297		Media Mentions over past 15 days	-0.00664	0.0154
	AOL (Mean)	-0.00853	0.0474				
	AOL (Std. Dev.)	0.0638	0.432	Time Trends	Yahoo × Day	-0.00231	0.00136+
					MSN × Day	0.00120	0.000711+
	Observations	99,494			Netscape × Day	-0.00144	0.000756+
	# Users[b]	300			Excite × Day	0.00348	0.00189+
	LL	-87,027			AOL × Day	0.00253	0.00123*

[a] Altavista is the base brand

[b] uses a random sample of 300 users in order to reduce the computational burden

+p<0.10; *p<0.05; **p<0.01

Figure 1

Figure 1a: Yahoo and Competitor Market Shares

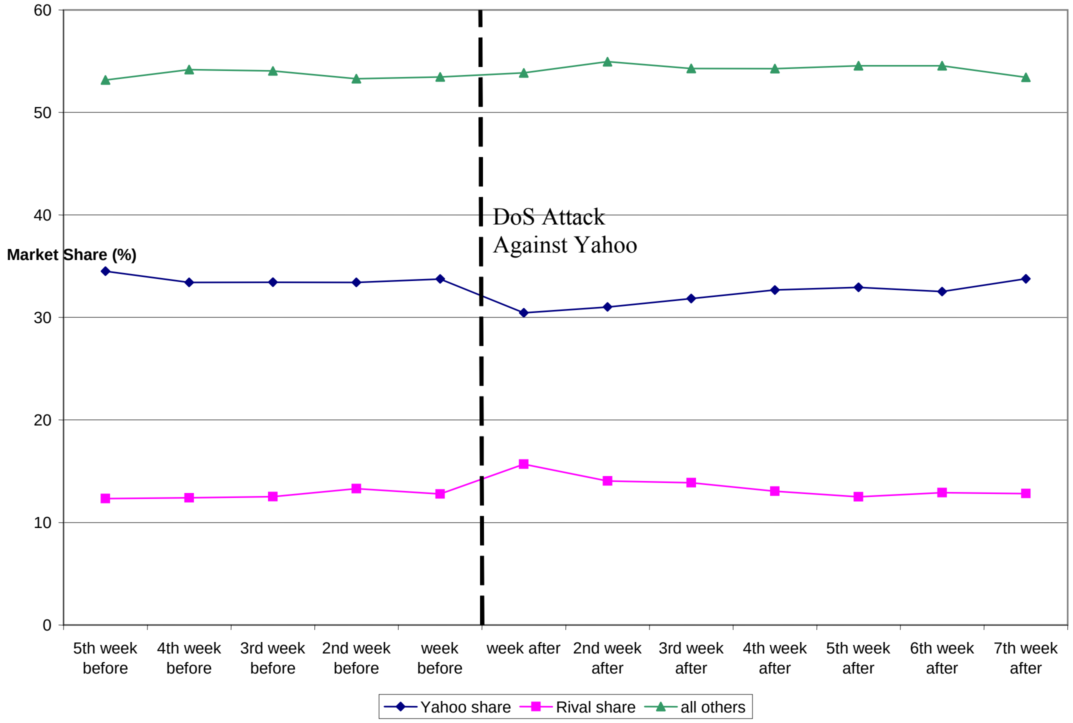


Figure 1b: CNN and Competitor Market Shares

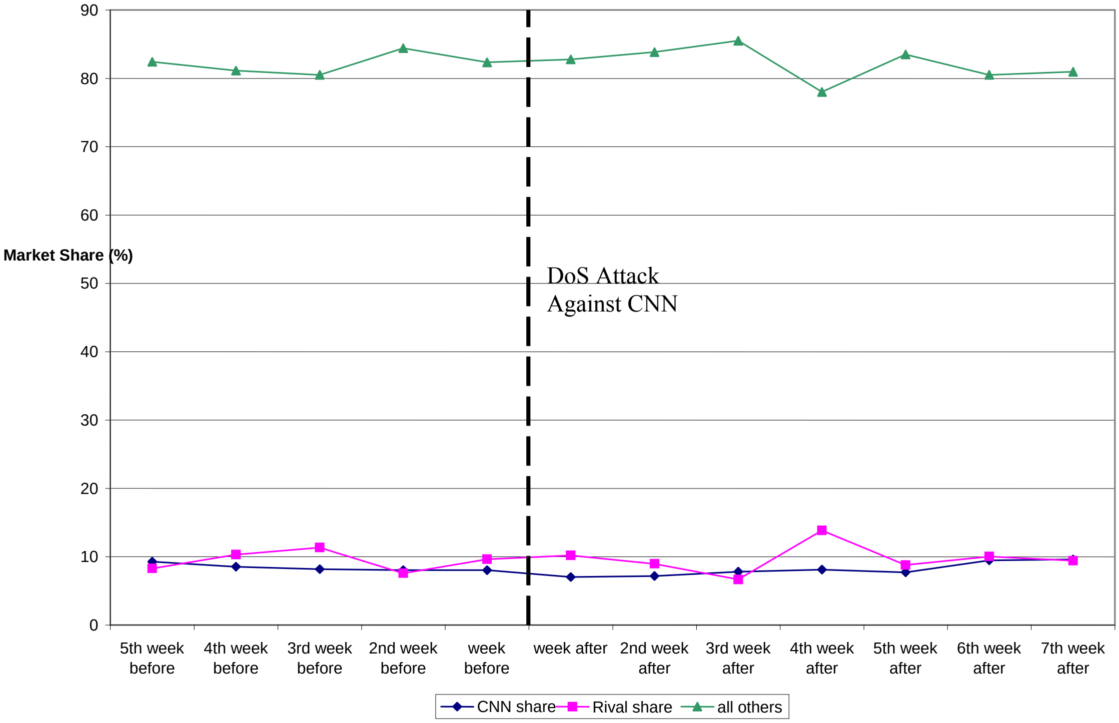


Figure 1c: Amazon and Competitor Market Shares

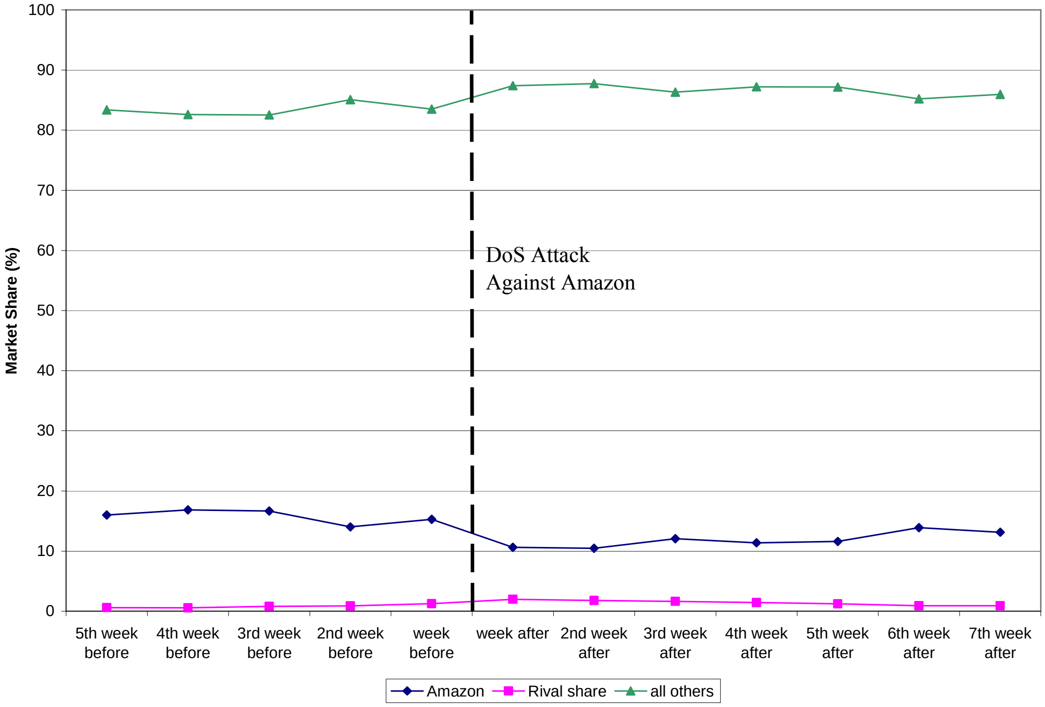
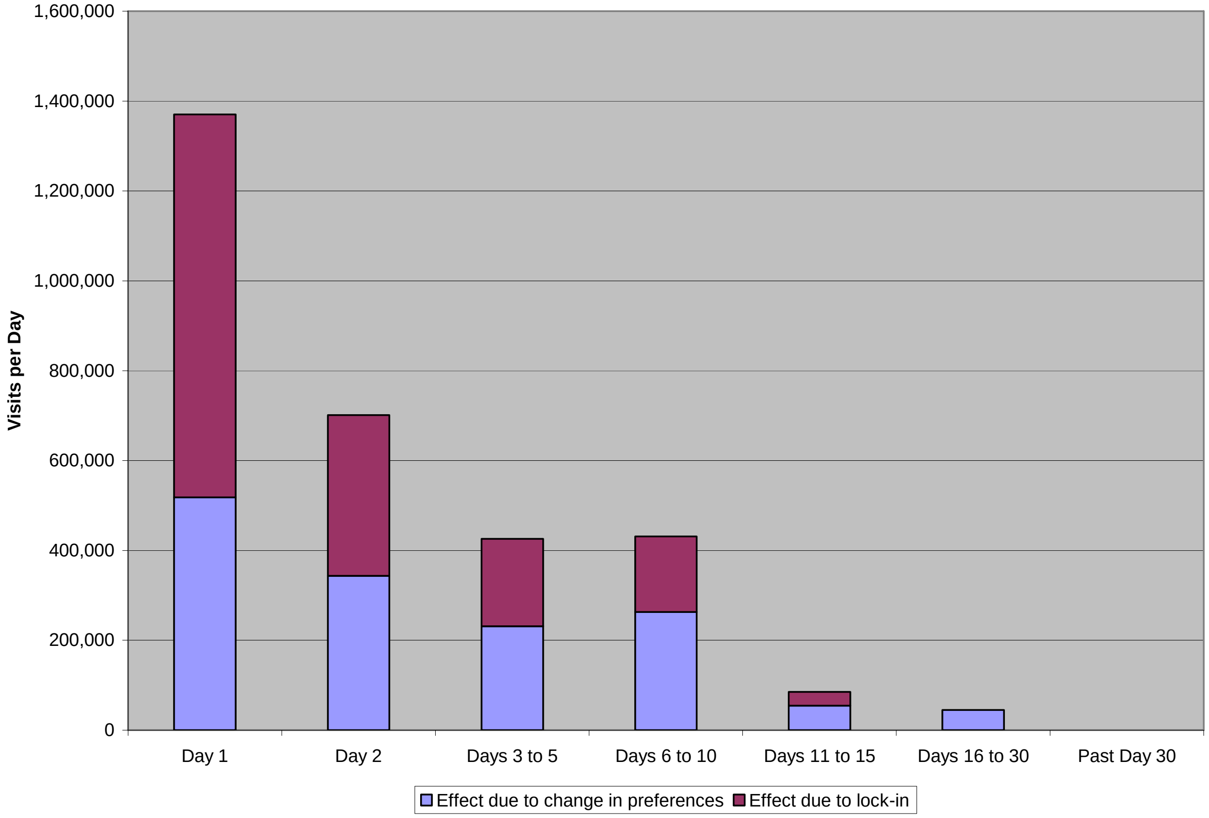


Figure 2. Estimated Effect of the Yahoo Attack on Rival Visits by Day (derived from Tables 7 and 8 column 1)



[*] University of Toronto, agoldfarb@rotman.utoronto.ca. I thank Mike Babyak of Plurimus Corporation for providing me with the data, the editor, two anonymous referees, Ulrich Kaiser, and seminar participants at SUNY Buffalo, Rutgers University, the ZEW conference on Information and Communication Technologies, the Fourth Workshop on the Economics of Information Security, and the BCRST conference for comments. This research was aided by SSHRC Grant #538-02-1013. Errors and omissions are my own.

[1] Mela, Gupta, and Lehmann (1997) define medium-term effects as effects that occur within 13 weeks of an event.

[2] Notable exceptions include Bell and Fitzsimons (1999), Jeuland (1979), and Swait and Erdem (2002). Bell and Fitzsimons look at the role of choice set size on utility from shopping at the retailer using both experimental data and scanner-panel data. Jeuland finds that consumers buy an alternative brand when their favorite brand in not available. Using scanner data, Swait and Erdem (2002) show that consistent availability leads to higher utility at the brand level. Rather than examining the impact of one incidence of unavailability on a brand, they explore how brands that are *often* unavailable generate lower utility.

[3] Also visits are not the relevant metric for Amazon. Amazon depends on purchases for revenue. Manchanda et al. (2006) show substantial differences between visits and purchases as a function of banner advertising exposure.

[4] This two-step method is admittedly inefficient relative to a model that simultaneously estimates the probability of being in the treatment group and the effect of the DoS attacks. Still, the large amount of data and the significance of the results suggest this inefficiency is not important for understanding the underlying processes in the data.

[5] The use of actual visits to other websites during the attack introduces another potential source of bias in the treatment group definitions. This may arise if the DoS attacks slowed down general access to the web. If this occurred the overall number of visitors may be pushed downward, biasing the results toward finding no effect of the attacks.

[6] Throughout the paper, I refer to the websites visited during the attacks as "rival" websites. Other competitors are labeled "other competing" websites.

[7] It is split into 1 day, 2 days, 3 to 5 days, 6 to 10 days, 11 to 15 days, 16 to 30 days, and over 30 days after the DoS attack. The spline allows for considerable flexibility in measuring the rate of decay. A previous version of this paper modeled decay as linear, revealing similar qualitative results.

[8] As in all natural experiments, there exist potential confounds to the results. The difference-in-difference identification deals with these to the extent possible; however, if there is something driving people into the treatment group that is also causing them to change their behavior then the results will be due to spurious correlation.

[9] The general results are robust to defining the media mentions variable as a dummy for whether the website was mentioned in the media that day. Bytes downloaded from the website, pages viewed, and time spent had the same general effect as bytes uploaded to the website with less explanatory power.

[10] The spline format means it is not possible to extrapolate out of sample. Using a linear decay function, the effect on Yahoo completely dissipates after 91 days.

[11] Simulated from the data and based on an estimated 43.3 million online households in February 2000.

[12] An example of this framework is Guadagni and Little's (1983) loyalty measure.

[13] This result is consistent with Johnson, Bellman, and Lohse's (2003) concept of "cognitive lock-in". They emphasize learning as an essential driver of online lock-in (switching costs). Consistent with the results in Table 9, their argument implies that inexperienced users would be most affected by lock-in.

[14] In the multinomial logit model, I use view length rather than bytes uploaded because it improves the model fit. Repeated search proxies for whether a given search failed, and is measured by whether the previous visit to the portal was followed by a visit to another portal. The variable *missing* has no economic interpretation. It is used to address the fact that before a user is observed to visit a particular portal in the data, past view length and search failure will be missing for that portal.