Volume 7, Issue 4
November 2014
ISSN: 1946-1836

Journal of Information Systems Applied Research

In this issue:

The **Journal of Information Systems Applied Research** (JISAR) is a double-blind peer-reviewed academic journal published by **EDSIG,** the Education Special Interest Group of AITP, the Association of Information Technology Professionals (Chicago, Illinois). Publishing frequency is currently quarterly. The first date of publication is December 1, 2008.

JISAR is published online (http://jisar.org) in connection with CONISAR, the Conference on Information Systems Applied Research, which is also double-blind peer reviewed. Our sister publication, the Proceedings of CONISAR, features all papers, panels, workshops, and presentations from the conference. (http://conisar.org)

The journal acceptance review process involves a minimum of three double-blind peer reviews, where both the reviewer is not aware of the identities of the authors and the authors are not aware of the identities of the reviewers. The initial reviews happen before the conference. At that point papers are divided into award papers (top 15%), other journal papers (top 30%), unsettled papers, and non-journal papers. The unsettled papers are subjected to a second round of blind peer review to establish whether they will be accepted to the journal or not. Those papers that are deemed of sufficient quality are accepted for publication in the JISAR journal. Currently the target acceptance rate for the journal is about 40%.

Questions should be addressed to the editor at editor@jisar.org or the publisher at publisher@jisar.org.

2014 AITP Education Special Interest Group (EDSIG) Board of Directors

Journal of Information Systems Applied Research

Editors

JISAR Editorial Board

Online versus In-Store: Price Differentiation for Multi-Channel Retailers

Javier Flores
jflores6@utpa.edu

Jun Sun
jsun@utpa.edu

Computer Information Systems and Quantitative Methods,
The University of Texas–Pan American
Edinburg, Texas, 78539, USA

Abstract

Pure-play online retailers have created pressure on traditional bricks-and-mortar retailers forcing many of them to move to a multi-channel business model to provide customers online storefronts in addition to local physical stores. Conventional wisdom suggests that online prices be lower than in-store prices. This study investigates whether multi-channel retailers follow such clear-cut pricing strategies based on the sampling of both their online and in-store prices. The results from three national retailers of office supplies in the USA indicate that their online prices are not necessarily lower than in-store prices. In addition, they suggest that the retailers apply different pricing strategies across different product categories. The finding reveals that multi-channel retailers use differential strategies for online and in-store pricing. It calls for further investigation of the interaction between consumer behavior and pricing strategies in hybrid e-commerce environment.

Keywords: Price differentiation, Multi-channel retailers, Pure-play online retailers, Bricks-and-mortar retailers.

1. INTRODUCTION

The advances in information and communication technologies (ICTs) have significant impacts on the operation modes of retailing industry. Traditional bricks-and-mortar retailers face the challenges from pure-play online retailer (e.g. Amazon.com, eBay.com, and Netflix.com). The competition forces many traditional companies to establish online storefronts for their customers in addition to local physical stores. Such a move creates multi-channel retailers from which people can buy products both online and in-store.

The main incentive for pushing the online channel is to reduce cost through the optimization of inventory, elimination of unnecessary intermediaries, and enhancement of customer relationship management (Fleischmann, Hall, & Pyke, 2004). Moreover, this new channel can provide consumers with richer and more accessible information (Brynjolfsson & Smith, 2000).

Most existing studies focus on the comparison between traditional bricks-and-mortar retailers and pure-play online retailers (Stylianou, Kumar, & Robbins, 2005). However, many large chain retailers, such as Best Buy, Barnes & Noble, or Walmart, operate on a hybrid mode by offering customers both online and in-store channels (Bernstein, Song, & Zheng, 2008). The prices that these multi-channel retailers offer online and in-store may or may not be the same. Though some researchers suggest that online prices tend to be lower than in-store prices for multi-channel retailers (Ratchford, Pan, & Shankar, 2003; Pan, Ratchford, & Shankar, 2002), few empirical studies investigate the issue by actually comparing prices.

The examination of the phenomenon is important because retailers with mixed channels can have different pricing strategies leading to different consumer experiences and competitive advantages. ICTs provide customers with different options to interact with companies in their shopping experiences. For example, multi-channel customers are inclined to use more than one channel to interact with organizations, possibly using companies' websites to check for information and prices, but buying in physical stores (Rangaswamy & Van Bruggen, 2005).

Retailers continue innovating in the usage of ICTs to support distribution channel, although important issues still need to be studied, such as online consumers' psychology to better understand websites success and failures (Wareham, Zheng, & Straub, 2005). Companies still need more experience with internet since it remains a new environment (Berstein, Song, & Zheng, 2008). There is still a lack of agreement on the factors determining acceptance of websites (Flavian, Gurrea, & Orus, 2009). Few studies have examined price differences for the same product across different channels within the same retailer. The understanding of these possible price variations can provide a clearer view of the e-commerce evolution.

This study examines whether the prices in multi-channel retailers differ in their two channels, the physical and the online. In specific, we verify whether the price differences follow the same pattern among stores, and across groups of products. The contribution of this study is on the analysis of possible price strategies among retailers in the same business area, identifying price differences among groups of products and between the two channels within retail companies.

The rest of the paper follows with a literature review on price strategies for companies, price differences for conventional and online channels within multi-channel retailers, and variations on prices for groups of products based on their characteristics. Furthermore, we describe the methodology utilized for the analysis, and list the corresponding results. Later, we present the discussions on the results, with the associated conclusion and implications.

2. LITERATURE REVIEW

Few related studies have been performed for the retail sector, even though ICTs are considered strong influencers to most industries (Doherty & Ellis-Chadwick, 2006). Retailers' evolution to a multi-channel mode still holds many unsolved issues to study. For example, it is not clear for a multi-channel retailer whether the online prices are lower than in-store prices. Huang & Swaminathan (2009) mention an Ernst and Young survey where two-thirds of multi-channel companies price their products the same for their conventional and online channels; with the note that customers usually expect lower online prices.

Along the maturation of ICTs and the retailers' migration to a multi-channel mode, a company may give two different prices for a single product: an in-store price and an online price. Lee, Kauffman, & Bergen (2009) argue that online reputation and relative price levels influence prices for different product categories. Previous empirical studies either compare the price for the pure-play online retailers to the brick-&-mortar retailers, but few examine the difference between online prices and in-store prices within each multi-channel retailer.

Consumer preferences are based on the retail format and on the price wanted (Keen, Wetzels, de Ruyter, & Feinberg, 2004). Price differences can occur at company level, channel level, and at product category level (Smith, Bailey, & Brynjolfsson, 2001).

Pricing strategies at company level

The Internet gives the idea of a more dynamic environment where it is easier to change prices, although this view does not consider the internal cost for companies to communicate, educate and even convince staff. It neither, considers the associated cost with the retrain of the sales force to a different organizational structure, to a different selling model, or how to take full advantage of the new price strategy (Bergen, Ritson, Dutta, Levy, & Zbaracky, 2003). On the other hand, Stylianou, Kumar, & Robbins (2005) asserted that contrary to what could be expected, price changes are not more frequent or different in magnitude for the online channel than for the conventional channel. Even though, cost and price dispersion have shown higher for the online channel.

The effects of ICTs on the business arena seem to rely on the generally accepted belief that they imply a more dynamic, effortless, and highly efficient medium, which push prices to lower levels, eliminate unnecessary intermediaries, and where consumers benefit from information at their fingertips. Nonetheless, different researchers affirm that activities like price adjustments can imply large quantities of time and effort from companies. This, without considering managerial costs involving information gathering, decision-making, and communicating the changes (Zbaracki, Ritson, Levy, Dutta, & Bergen, 2004).

Price changes have a direct impact on operations, and vice versa. They can also have dramatic effects on supply chain. Researchers called for additional analysis on the relationships between dynamic pricing and inventory, production planning, and capacity management decisions (Fleischmann, Hall, & Pyke, 2004).

E-commerce eases dynamic pricing practices, where they are linked to groups and individual preferences (Haws & Bearden, 2006). Price difference between products from Internet companies becomes smaller as the number of companies competing increases (Baye, Morgan, & Scholten, 2004b). Besides, online price dynamism suggests the idea of effortless price changes, which does not take into account the implications such as consumer perceptions on fairness (Haws & Bearden, 2006) and feelings of discrimination by the dynamic pricing, having as consequence lost of trust (Kannan & Kopalle, 2001). In the same way, price changes are perceived unfair when they are done in a shorter period of time, especially with low priced products. Differences between consumers result in the greatest perceptions of unfairness and the lowest level of satisfaction. The highest perception of fairness and satisfaction across all price level conditions is reached when the consumers are involved to set prices.

All this defines price strategies that retailers establish to compete in a faster, and usually high responsive electronic environment. The Internet platform allows having mixed pricing strategy, where online retailers with higher quality in their services can benefit from a competitive market. This higher quality can help to differentiate them, and be able to set higher prices, using also the obtained trust and reputation (Venkatesan, Mehta, & Bapna, 2006).

Part of these price strategies might involve random price changes to hinder customers' learning from low price practices (Varian, 1980). Baye, Morgan, & Scholten (2004a) provide the 'hit and run' sale as a way to avoid getting into competition where the minimum price is forced. Here, in order to maximize profits, online retailers need to be as much unpredictable as possible, changing timing and discount magnitudes. It is hard, if not impossible for a customer to learn from low-prices when store have continuous differences among prices (Lach, 2002).

Thus, differences among retailers can be significant, influencing our study. We considered three different companies, all of them being national retailers of office supplies with a solid presence through physical and online locations.

Pricing strategies for business modes

Many researchers agree on the idea that prices for the pure-play online retailers should be lower than those of conventional retailers. Ancari & Shankar (2002) argued that conventional retailers have the highest prices, followed by the multi-channel retailers, and ending with pure-play online retailers with the lowest price. However, they also suggested that when shipping costs are included, the order change, from multi-channel retailers having the highest price, to pure-play online retailers, and ending with conventional retailers showing the lowest price. Nevertheless, it is a common practice nowadays that most multi-channel retailers offer ship-to-store services for free. Many customers place orders online and pick them up in store. This gives multi-channel retailers further advantage in terms of both customer convenience and cost saving.

Additionally, Ratchford, Pan, and Shankar (2003) posit that ICTs can improve consumer position in the buying process, and that online prices are usually lower than prices in traditional channels. At the same time, different types of customers have different online buying preferences. For example, goal oriented buyers look for efficient and strong economic value options, whereas experiential buyers prefer enjoyable purchasing experience (Mathwick, Malhorta, & Rigdon, 2002).

As mentioned earlier, researchers postulate that online retailers present lower prices than traditional ones (Brynjolfsson & Smith, 2000; Pan, Ratchford, & Shankar, 2002). Bailey (1998) argues that Internet retailers show different prices suggesting diverse strategies, including those for homogenous products. Where, those companies are possibly contending for the same groups of customers, although using different marketing practices and different kind of services.

Most of the retailers have a presence in the Internet showing a new paradigm, nevertheless, challenges such as channel conflicts still need to be cleared (Webb, 2002). Schoenbachler & Gordon (2002) argue that marketers working with multi-channel face problems such as the cannibalization of sales with higher margins. In addition, they deal with high costs to implement campaigns, and issues related to customers retention. At the same time, studies also show that channel conflicts can be reduced by combining the advertisement and general information transmission process, together with adapting prices for both channels (Zhang, Zhuang, & Huang, 2010). This equilibrium price is possible for conventional and online channels, where usually it is close to the price of conventional channel considering that the online price is lower (Yao & Liu, 2005).

All three retailers which were considered for our study have both channels, the conventional using physical stores, and the online through their websites. Particularly, all three retailers share same markets with similar groups of products, having a strong physical presence in the south area of the state of Texas.

Pricing strategies for product nature

Bock, Lee, & Li (2007) studied price differences among online retailers, suggesting that differences in Internet maturity have an impact on retailers. Particularly, they compared retailers from the United Stated and China, where US retailers have lower price dispersion. Findings show that price levels change depending on the product types regardless of the Internet maturity; and that online retailers usually have lower prices and lower price differences compared to multi-channel retailers.

Furthermore, the price dispersion among online retailers is linked to service characteristics which can allow higher prices. Multi-channel retailers with established brands in physical stores can better manage price premiums compared to pure-play online retailers, although this is not observed in high competitive markets such as books, CDs, and flight tickets (Walter, Gupta, & Su, 2006). In addition, hybrid retailers can be more successful than pure-plays due to advantages in brand strength, cross-promotional opportunities, and the multi-channel offering (Min & Wolfinbarger, 2005).

Studies posit that there is a potential to replace the traditional channel with e-commerce for complex or technological products, which have no standard characteristics (Jantan, Ndubisi, & Yean, 2003). In contrast, for some types of retailers, their customers assign higher value to physically displayed products at their stores as consequence of the possibility to prove them personally. However, after the customer bought the product once, they tended to have the same product valuation through the conventional channel as online. For this reason, the retailer is inclined to set prices to attract customers to the physical store initially, and then take advantage of the increased profits from online sales (Mehra, Kumar, & Raju, 2010).

The reasons for us to include different groups of products were based on these previous facts in the literature. Particularly, around differences on customers' preference, and differences on prices for varied products assigned accordingly to their characteristics. From the simplest ones with not much differentiation (e.g. paper, envelopes) to the more complex such as electronic products, which can still cross-sell related services from retailers.

In summary, companies can define different price strategies which may vary across different channels and even across different product groups. To enhance the comprehension on these events, we did a case study with three national retailers of office supplies, all of which have a conventional channel plus an online channel. Six groups of products were selected considering their characteristics that could influence price differences.

3. METHODOLOGY

As previously mentioned, the goal of this study is to analyze price differences in multi-channel retailers, across groups of products. For which it is preferred that the chosen companies be all national multi-channel retailers with a strong local presence.

The office supply market is highly competitive as the sales of products are highly price sensitive.

Thus, this study samples both online and in-store prices from major office supply retailers in USA, including OfficeMax, Office Depot and Staples. In order to avoid bias in the data and retain confidentiality of the results, they are labeled as Store A, Store B and Store C after a scrambling of the order.

We collected and compared prices for different groups of products within the three office supplies retailers. These groups included Machine Supplies (MS), Office Technology (OT), Filing and Storage (FS), Paper (PA), Personal Organizers (PO), and Desktop Accessories (DA). The variety of products show not only price differences, but distinguishable characteristics, such as for printers that are not only much more complex than a ream of paper, but which can imply additional sales through attached services and warranties extensions.

The three retailers showed products in all six categories, and had a well-developed web site. Their physical stores are located in highly transited avenues, and have a continuous customer flow, besides their direct sales to medium and large size organizations.

One of the authors went to the physical stores to record the prices of different products. For each group of products, at least 30 prices were collected. This ensures sufficient sample size for the statistical analyses including *t*-test and ANOVA *F*-test to compare mean prices. During the same period of time, the research team search for the online prices of the same products from each retailer. All the prices recorded and used in this study were regular ones, and special, clearance and on-sale prices were not included. This avoids the potential bias due to special events such as sales and promotions.

There is a wide diversity in the prices of different products, such as the price for a pencil compared to a digital camera. To make the comparison, we used price ratios created dividing the online price of a product by its in-store price. In this way, the comparison is based on the changes in percentages, providing a better perspective of variations.

4. RESULTS

Table 1 compares the online prices and in-store prices in terms of their price ratios. If the prices are the same, the ratio is 100%. A lower-than-100% ratio indicates that the average online price is lower than the average in-store price, and a higher-than-100% ratio indicates that the average online price is higher than the average in-store prices. The overall average price ratio for all the products in the sample is 98.92% and it is not significantly different from 100%.

	Store			
Product	A(n=89)	B(n=131)	C(n=78)	By Prod
MS (n=72)	100.00 (.00)	99.17 (3.33)	101.25 (6.87)	99.96 (3.96)
OT (n=36)	97.22 (9.62)	100.00 (.00)	104.49 (15.55)	100.57 (10.69)
FS (n=59)	100.77 (2.43)	90.76*** (18.52)	99.94 (10.18)	94.17*** (16.09)
PA (n=32)	100.00 (.00)	92.96 (27.85)	106.40*** (6.19)	99.36 (17.84)
PO (n=49)	100.55 (2.06)	96.09 (12.51)	107.60* (14.39)	99.95 (11.84)
DA (n=50)	100.00 (.00)	103.13 (12.53)	98.63 (13.88)	100.54 (10.70)
By Store	99.80 (3.73)	96.16*** (15.18)	102.55* (11.84)	98.92 (12.18)

Table 1: Online Price to In-store Price Ratio (%)

Note: Standard deviations are given in parentheses below the mean. *-Significant at 0.1 level, **-Significant at 0.05 level, ***-Significant at 0.01 level.

Among the three stores, however, Store B offered relatively lower online prices than in-store prices by almost 4% on average, but Store C makes their online prices higher on average than their in-store prices by less than 3%. Store A, on the other hand, did not have significantly different on-line and in-store prices as the average price ratio is very close to 100%.

Category-wise, only the Filing & Storage category had significantly lower online prices than in-store prices by an average of 5.83% (i.e. 100% - 94.17%). The online and in-store prices for the other five categories were not much different.

Nevertheless, Store C had the categories of Paper and Personal Organizers with significant differences between online and in-store prices with 106.40% and 107.60% respectively. Moreover, Store B showed one price ratio with significant difference between online price than in-store price for the Filing and Storage category with a 90.76% of the online prices lower than in-store prices, on average.

Table 2 gives the ANOVA results for testing mean differences in the price ratios across three stores for all the products and each product category respectively. There was a significant difference in the price ratio between online and in-store prices at a .01 level across stores for all product categories.

Product	*F*	Significant Paired Difference
MS	1.51	C-B: 2.08%*
OT	1.45	C-A: 7.27%*
FS	2.53*	A-B: 10.01%*; C-B: 9.18%*
PA	1.62	C-B: 13.44%*
PO	4.05**	C-B: 11.51%***
DA	.75	
Overall	7.37***	C-B: 6.39%***; A-B:3.64%**

Table 2: Store-wise ANOVA Tests

Note: H0: $\mu_A = \mu_B = \mu_C$; *-Significant at 0.1 level; **-Significant at 0.05 level; ***- Significant at 0.01 level.

The post-hoc examination located two pairs that are significantly different: Store B - Store A and Store B – Store C. That indicated that Store B is different from both Store A and Store C in terms of price ratios, but Store A and Store C are not that different. Thus, the stores can be divided into two groups: Store B by its own in one group and Store A and Store C in another group. The difference in price ratio between Store B and Store C was 6.39%. As shown in table 1, Store B offered lower online prices than in-store prices by 3.84% (i.e. 100%-96.16%) on average, but Store C made the online prices higher than in-store prices by an average of 2.55% (i.e. 102.55%-100%). Thus, the total gap of 6.39% between the two stores can be decomposed into 3.84% plus 2.55%. In the same way, the difference in price ratio between Store B and Store A was 3.64%. As shown in table 1, Store B offered lower online prices than in-store prices by 3.84% (i.e. 100%-96.16%) on average, but Store A had the online prices slightly below in-store prices by an average of 0.20% (i.e. 100%-99.80%). Thus, the total gap of 3.64% between two the stores can be decomposed into 3.84% minus 0.20%.

For the Machine Supplies products three groups of stores can be set up, starting with the lower-price-ratio group which includes the Store B with a price ratio of 99.17%. Additionally, there was the higher-price-ratio group comprised by Store C, having a price ratio of 101.25%, and establishing a significant difference to the lower-ratio group by 2.08%. In the middle, Store A is shown with a 100% price ratio, not having a significant difference with neither of the other two stores.

A similar situation occurred for Office Technology products where the lower-price-ratio group comprised by Store A with a price ratio of 97.22%, and the higher-price-ratio group, which includes only Store C (104.49%) showed a significant difference totaling 7.27%. In the middle, Store B displayed a price ratio of 100%, having no significant difference to Store A neither to Store C.

Filing & Storage products displayed in the post-hoc examination two groups of stores, starting with the lower-price-ratio comprised by Store B only with a price ratio of 90.76%, and the higher-price-ratio group including Store A (100.77%) and Store C (99.95%). The significant differences between stores in each group were about 9.5%. For Paper products, the lower-price-ratio group included only Store B with a price ratio of 92.96%; compared to the higher-price-ratio group comprised by Store C only, with a price ratio of 106.40%, displayed a significant difference of 13.44%. In the middle, Store A had a price ratio of 100%, pointing no significant difference to the other two groups.

For Personal Organizers products a similar situation to the Paper category was shown, with the lower-price-ratio group comprised by Store B, having a price ratio of 96.09%, and the higher-price-ratio group including Store C with a price ratio of 107.60%. A significant difference between these two groups was estimated, totaling 11.51%. In the middle, Store A had a price ratio of 100.55%, without any significant difference to the other two groups. Finally, Desktop Accessories products showed no significant differences among the three stores: Store B (103.13%), Store A (100%), and Store C (98.63%).

Table 3 gives the ANOVA results for testing mean differences in the price ratios across six product categories for all the stores and each store respectively. Cross product categories for all the stores had a significant difference in the price ratio between online and in-store prices at a .05 level. The category-wise comparison in Table 1 shows that the online prices are on average 94.17% of the in-store prices for Filing and Storage products, whereas the differences are not that significant for other categories.

The post-hoc examination located diverse pairs of categories with significant differences, which we grouped to distinguish categories. For all the stores within the overall, two groups can be created: the lower-price-ratio group comprising the Filing & Storage included, listing the lowest online price to in-store price ratio (94.17%). Subsequently, the high-price-ratio group with the rest of the categories: MS (99.96%), OT (100.57%), Paper (99.36%), PO (99.95%), DA (100.54%). The average significant difference in price ratios between the two groups was about 6%.

Store	*F*	Significant Paired Difference
A	1.46	MS-OT: 2.78%**; FS-OT: 3.55%**; PA-OT: 2.78%*; PO-OT: 3.33%**; DA-OT: 2.78%**
B	2.23*	MS-FS: 8.41%**; OT-FS: 9.24%*; DA-FS: 12.38%***; DA-PA: 10.17%*
C	1.21	PA-DA: 7.77%*; PO-DA: 8.97%**
Over-all	2.34*	MS-FS: 5.80%***; OT-FS: 6.41%***; PA-FS: 5.19%**; PO-FS: 5.78%**; DA-FS: 6.37%***

Table 3: Category-wise ANOVA

Note: H0: $\mu_{MS}=\mu_{OT}= \mu_{FS} = \mu_{PA} = \mu_{PO}= \mu_{DA}$; MS - Machine Supplies; OT - Office Technology; FS - Filing & Storage; PA – Paper; PO - Personal Organizers; DA - Desktop Accessories; *-Significant at 0.1 level, **-Significant at 0.05 level, ***- Significant at 0.01 level.

Store B was the only store to show an overall significant difference (p-value =0.055) for the price ratio at store level. The post-hoc analysis identified two groups: the lower ratio group comprised the Filing and Storage products and Paper products (90.76% and 92.96% respectively) and the higher ratio group comprised Machines Supplies, Office Technology, and Desktop Accessories products (99.17%, 100%, and 103.13% respectively). The average differences in price ratios between two groups were about 10%. In the middle laid the Personal Organizer category that was not significantly different from either group (i.e. 96.09%).

Even though Store A did not display an overall significant difference at store level (p-value=0.212), the post-hoc examination displayed the Office Technology products with significantly difference to the rest of the categories due to its lower-price-ratio of 97.22% that suggests a lower online price than in-store price. Consequently, the lower price group included only the OT category, and the higher price group comprised all the rest of the categories: MS (100%), FS (100.77%), Paper (100%), PO (100.55%), and DA (100%). The average differences in price ratios between these two groups were about 3%. On the other hand, Store C had no overall significant difference at store level (p-value=0.312), showing in the post-hoc analysis that the Desktop Accessories products were significantly different to Paper and Personal Organizers categories, creating the lower price group with the Desktop Accessories products only (98.63%), and higher price group with Paper and Personal Organizers categories (106.40% and 107.60% respectively). The average significant differences in price ratios between group 1 and group 2 were about 8%. In the middle could be found the other three categories: MS (101.25%), OT (104.49%), and FS (99.94%) without any significant difference to either group.

5. CONCLUSION AND IMPLICATIONS

Retailers are experiencing changes on their business models adding an online channel as an option to better reach their customers. Many companies in the area of office supply are using a multi-channel model having their conventional bricks-and-mortar stores plus their websites supporting an online store.

The conventional wisdom indicates that retailers offer lower prices through their online offerings; although most studies have focused comparing pure-play retailers to traditional bricks-and-mortar, not having a clear idea if in-store prices are really above the online prices within the same retailer.

Consumers are learning to deal with this new model, sometimes checking prices and product information online and buying the product in the physical stores. Even requesting additional details on products via ICTs, and sending their opinions and preferences back to the retailer electronically. Companies using this new medium still need additional experience to fully take advantage of it, moving their prices across groups of products, channels, and facing a more elaborated competition from other retailers.

Differences in prices along these dimensions (stores, channels, product categories) are assumed not purely random, but the result of pricing strategies. Three stores were selected, Store A, Store B, and Store C, all national retailers of office supplies, with a solid presence in this highly competitive market. The three of them have both channels bricks-and-mortars stores, and an online presence. All of them have products in all selected product categories (MS-Machine Supplies; OT-Office Technology; FS-

Filing & Storage; PO-Personal Organizers; DA-Desktop Accessories, and Paper).

In order to analyze these differences we compared price ratios obtained by dividing the online price by the in-store price. This was needed since the product categories included products with different price magnitudes, misleading the results, and a price ratio provides a percentage amount easier to compare.

Results suggest an overall difference between online and in-store prices not significant, which goes against the generally accepted beliefs and some previous studies. Table 1 shows the overall price ratio for all the products in the sample, 98.92%, presenting a minor difference on average for the online prices to the in-store prices across stores and categories.

Different stores have different pricing strategies in terms of how to differentiate online prices and in-store prices. Results showed that, Store C pushes the online prices higher than in-store prices. A possible explanation could be that they offer discounts in stores rather than online looking to keep the online prices consistent with the catalog prices. Another reason may include staff performance evaluation based on local sales volume, supporting in-store sales.

Nevertheless, as can be seen in Table 1, the three stores showed different product strategies where Store A displayed the same or very close online prices than in-store prices through the different categories. Store B, on the other hand, displayed lower online prices than in-store prices in four out of six categories, and at the store level also. In contrast, Store C had four categories with price ratios above 100% indicating higher online prices than in-store prices. Overall, and at four categories out of six, Store B showed a significant different price ratio than Store A and Store C, being the store with more differences in price ratios overall and across categories.

Across different product categories, stores employ different pricing strategies. In particular, Filing & Storage seems to have a lower price ratio than others. Even though, only the Filing & Storage and Personal Organizers categories in Table 2 indicated significant different means of price ratios among stores, overall the store-wise ANOVA test showed a significant different means of price ratios, pointing different price strategies for the three stores.

Each store would use different pricing strategies for different product categories. Some may give online discount, if any, for one product category at a time, while other may not. Table 3 shows test results for Store C and Store A as the two stores with a consistent price strategy across categories. For Store C, the price ratios were about the same except for Desktop Accessories, and for Store A the Office Technology category was the only one different to each and every one of the rest of categories. The overall result shows that the mean of price ratios across categories are different, indicating a diversified pricing strategy across products.

This study as all studies, have limitations, including a cross-sectional analysis where we were not able to track the price changes over a period of time. A longitudinal analysis can provide a better insight on the price strategies and associated consumer behavior. At the same time, it can give a more detail idea of past strategies used by companies looking to attract more customers, especially, more sales with ideally better overall profits.

The importance of this study relies on the fact that these price differences across channels and product categories affect consumer experiences and their behaviors. Retailers can attract different consumer groups depending on their preferences (traditional, object-driven, experience-driven). A better understanding of this commercial interaction can help to support decisions toward optimizing sales in the retail sector.

6. REFERENCES

Ancari, F., & Shankar, V. (2004). Price levels and price dispersion within and across multiple retailer types: Further evidence and extension. *Journal of the Academy of Marketing Science, 32*(2), 176-187. DOI 10.1177/0092070303261464

Baye, M. R., Morgan, J., & Scholten, P. (2004a). Temporal price dispersion: Evidence from an online consumer electronics market. *Journal of Interactive Marketing, 18*(4). DOI 10.1002/dir.20016

Baye, M. R., Morgan, J., & Scholten, P. (2004b). Price dispersion in the small and in the large: Evidence from an internet price comparison. *The Journal of Industrial Economics, 52*(4), pp. 463-496.

Bayle, J. P. (1998). Intermediation and electronic markets: Aggregation and pricing in internet commerce. Ph.D. dissertation, Massachusetts Institute of Technology, 1998.

Bergen M., Ritson, M., Dutta, S., Levy, D., & Zbaracky (2003). Shattering the myth of costless price changes. *European Management Journal, 21*(6), pp.663-669.

Bernstein, F., Song, J., & Zheng, X. (2008). Bricks-and-mortar vs. clicks-and-mortar: An equilibrium analysis. *European Journal of Operational Research, 187*, pp. 671-690. DOI 10.1016/j.ejor.2006.04.047

Bock, G., Lee, S. T., & Li, H.Y. (2007). Price comparison and price dispersion: Products and retailers at different internet maturity stages. *International Journal of Electronic Commerce, 11*(4), pp. 101-124. DOI 10.2753/JEC1086-4415110404

Brynjolfsson, E., & Smith, M. D. (2000). Frictionless commerce? A comparison of internet and conventional retailers. *Management Science, 46*(4), pp. 563-585.

Doherty, N. F., & Ellis-Chadwick, F. (2006). New perspectives in internet retailing: a review and strategic critique of the field. *International Journal of Retail and Distribution Management, 34*(4/5), 411-428.

Flavian, C., Gurrea, R., & Orus, C. (2009). A heuristic evaluation of websites design for achieving the web success. *International Journal of Services and Standards, 5*(1), 17-41.

Fleischmann, M., Hall, J. M., & Pyke, D. F. (2004). Smart pricing: Linking pricing decisions with operational insights. *MIT Sloan Management Review, 45*, 9-13.

Haws, K. L., & Bearden, W. O. (2006). Dynamic pricing and consumer fairness perceptions. *Journal of Consumer Research, 33*.

Huang, W. & Swaminathan, J. M. (2009). Introduction of a second channel: Implications for pricing and profits. *European Journal of Operational Research, 194*, 258-279.

Jantan, M., Ndubisi, N. O., & Yean, O. B. (2003). Viability of e-commerce as an alternative distribution channel. *Logistics Information Management, 16*(6), 427-439. DOI 10.1108/09576050310503402

Kannan, P. K. & Kopalle, P. K. (2001). Dynamic pricing on the internet: Importance and implications for consumer behavior. *International Journal of Electronic Commerce, 5*(3), 63-83.

Keen, C., Wetzels, M., de Ruyter, K., & Feinberg, R. (2004). E-tailers versus retailers which factors determine consumer preferences. *Journal of Business Research, 57*, 685-695.

Lach S. (2002). Existence and persistence of price dispersion: an empirical analysis. *NBER Working Paper Series*. Retrieved August 10, 2011 from http://www.nber.org/papers/w8737

Lee, D., Kauffman, R. J., & Bergen, M. E. (2009). Image effects and rational inattention in Internet-based selling. *International Journal of Electronic Commerce*, *13*(4), 127-166.

Mathwick, C., Malhotra, N. K., & Rigdon, E. (2002). The effect of dynamic retail experiences on experimental perceptions of value: an Internet and catalog comparison. *Journal of Retailing, 78*, 51-60.

Mehra, A., Kumar, S., & Raju, J. S. (2010). Relationship retailing using online and store channels. *20th Workshop on Information Technologies and Systems*, December 11-12, 2010. St. Louis, MO, USA.

Min, S. & Wolfinbarger, M. (2005). Market share, profit margin, and marketing efficiency of early movers, bricks and clicks, and specialists in e-commerce. *Journal of Business Research, 58*, 1030-1039. DOI 10.1016/jbusres.2004.02.005

Pan, X., Ratchford, B. T., & Shankar, V. (2002). Can price dispersion in online markets be explained by differences in e-tailer service quality?. *Journal of the Academy of Marketing Science, 30(*4), 433-445. DOI 10.1177/009207002236915

Rangaswamy, A., & Van Bruggen, G. H. (2005). Opportunities and challenges in multichannel marketing: An introduction to the special issue. *Journal of interactive marketing, 19*(2), 5-11. DOI 10.1002/dir.20037

Ratchford, B. T., Pan, X., & Shankar, V. (2003). On the efficiency of internet markets for consumer goods. *Journal of Public Policy & Marketing, 22*(1), 4-16.

Schoenbachler, D. D. & Gordon, G. L. (2002). Multi-channel shopping: understanding what drives channel choice. *Journal of Consumer Marketing, 19*(1), 42-53. DOI 10.1108/07363760210414943

Smith, M., Bailey, J., & Brynjolfsson, E. (2001). Understanding digital markets: Review and assesment. Retrieved from http://www.econ2.jhu.edu/People/Harrington/375/sbb99.PDF

Stylianou, A., Kumar, R.L., & Robbins, S.S. (2005). Pricing on the internet and in conventional retail channels: A study of over-the-counter pharmaceutical products. *International Journal of Electronic Commerce, 10*(1), 135-148.

Varian, H. R. (1980). A model of sales. *The American Economic Review, 70*(4), 651-659.

Venkatesan, R., Mehta, K., & Bapna, R. (2006). Understanding the confluence of retailer characteristics, market characteristics and online pricing strategies. *Decision Support Systems, 42*, 1759-1775.

Walter, Z., Gupta, A., & Su, B. (2006). The source on on-line price dispersion across product types: An integrative view of on-line search costs and price premiums. *International Journal of Electronic Commerce, 11*(1), 37-62. DOI 10.2753/JEC1086-4415110102

Wareham, J., Zheng, J. G., & Straub, D. (2005). Critical themes in electronic commerce research: a meta-analysis. *Journal of Information Technology, 20*, 1-19. DOI 10.1057/palgrave.jit.2000034

Webb, K. L. (2002). Managing channels of distribution in the age of electronic commerce. *Industrial Marketing Management, 31*, 95-102.

Yao, D. Q. & Liu, J. J. (2005). Competitive pricing of mixed retail and e-tail distribution channels. *OMEGA, 33*, 234-247.

Zbaracki, M., Ritson, M., Levy, D., Dutta, S., & Bergen, M. (2004). Managerial and customer costs of price adjustment: Direct evidence from industrial markets. *The Review of Economics and Statistics, 86*(2), 514-533.

Zhang, T., Zhuang, G., & Huang, Y. (2010). A literature review on the studies of internet retailing management. *Technology and Investment*, *1*, 19-25

Editor's Note:

This paper was selected for inclusion in the journal as a CONISAR 2013 Meritorious Paper. The acceptance rate is typically 15% for this category of paper based on blind reviews from six or more peers including three or more former best papers authors who did not submit a paper in 2013.

Similarity and Ties in Social Networks: a Study of the YouTube Social Network

Amir Afrasiabi Rad
a.afrasiabi@uOttawa,ca
University of Ottawa
Ottawa ON K1N 6N5, Canada

Morad Benyoucef
Benyoucef@Telfer.uOttawa.ca
Telfer School of Management
University of Ottawa
Ottawa, ON K1N 6N5, Canada

Abstract

Social networks and the propagation of content within social networks have received an extensive attention during the past few years. Social network content propagation is believed to depend on the similarity of users as well as on the existence of friends in the social network. Our former investigation of the YouTube social network showed that strangers (non-friends and non-followers) play a more important role in content propagation than friends. In this paper, we analyze user communities within the YouTube social network and apply various similarity measures on them. We investigate the degree of similarity in communities versus the entire social network. We found that communities are formed from similar users. At the same time, we found that there are no large similarity values between friends in YouTube communities.

Keywords: Social Network Analysis; Similarity; Social Ties; Influence; YouTube

1. INTRODUCTION

Social networking websites, such as MySpace, Facebook, Twitter, Flickr, Orkut, YouTube, etc. are becoming more and more popular. Statistics show that in the US, almost 90% of the teenage and young adult age group are social network users (Trusov, Bodapati, & Bucklin, 2010). The birth of Web 2.0 allowed Internet content users to become Internet content providers as well. Social networks, as Web 2.0 applications, contribute their share to this paradigm shift. Social network users upload more than 35 hours of videos to YouTube every minute (YouTube LLC., 2010); and they contribute to Facebook by generating more than 30 billion pieces of content when they spend over 23 billion minutes on Facebook every month (Facebook Inc., 2011). Also, a billion tweets every month (Twitter Inc., 2011) is another indicator of this paradigm shift. Hence social networks are turning into hubs of social activity. Along with their popularity as a new communication medium, social networks are regarded as tools for social presence and for building social identity (Rad, Amir, & Benyoucef, 2011). The interconnected nature of social networks is a building block for establishing social identity. This is because social identity has no meaning if it is not defined in the context of a society. Social identity is always accompanied with ideas, or user generated content in the context of a social network. These ideas get propagated through interconnections between people in a social network, and they work as a way of further establishing social identity. Therefore, it is the interconnectivity of users in online social networks that allows user generated content, ideas, and influence to be easily propagated through the whole social network (Afrasiabi Rad & Benyoucef, 2012).

The wide use of social networks and their ability for propagating ideas attracted the attention of the marketing community which soon realized that content propagation along social links can lead to a huge community of users who can be used as viral advertisers. Moreover, the unique characteristics of social networks provide the opportunity to harness the collective opinions of the population in order to shape user behavior through adequate marketing campaigns while gaining insights into current and future market trends (Asur & Huberman, 2010; Bearden, Calcich, Netemeyer, & Teel, 1986; Leskovec, Adamic, & Huberman, 2007). There has been numerous studies on the different aspects, enablers, and contributing factors of viral advertisement on social networks (Bearden et al., 1986; Domingos & Richardson, 2001; Duan, Gu, & Whinston, 2008; Evans, 2009; Hu, Tian, Liu, Liang, & Gao, 2011; Kempe, Kleinberg, & Tardos, 2005; Kim & Srivastava, 2007; Stephen & Toubia, 2009; Van den Bulte & Joshi, 2007). However, there is little research dedicated to discovering why and how idea propagation occurs in the online world.

In one of our earlier studies, and in an attempt to analyze propagation, its characteristics, and its contributing factors, we investigated the propagation of data in an open social network (i.e., YouTube) (Afrasiabi Rad & Benyoucef, 2012). We define an "open" social network as a social network where privacy settings allow for content posted by a user to be seen by all members of a social network. In other words, privacy settings do not restrict viewing, commenting on, or sharing content to only friends or followers (also called subscribers on certain social networks such as YouTube) of a user. Based on our definition, social networks such as YouTube, Twitter and Flickr fall into the category of open social networks. Our previous study (Afrasiabi Rad & Benyoucef, 2012) revealed that content propagation in online open social networks follows different patterns compared to what has been observed in offline social networks (i.e., pre-internet social networks) (Judea, 1986). Although the actions of individuals are usually open to a wide range of other users in both offline and online open social networks, interestingly, propagation in offline social networks is mostly affected by the number of ties (i.e., friends, coworkers, and family) and their networks, while our study revealed that in an online open social network, propagation is far more affected by individuals who are neither in the network of friends nor the network of followers of the content generator.

Other studies also revealed contradictory results. For instance, Crandall et al. (Crandall, Cosley, Huttenlocher, Kleinberg, & Suri, 2008) studied multiple online and offline social networks and discovered that an increase in similarity between online social network users boosts both the magnitude and speed of content propagation. On the other hand, and focusing merely on offline social networks, Feld (Feld, 1981) discovered that similarity is one of the major factors that define the strength of ties between members of a social network. Note that in this paper, we use "ties", "links", "connections" and "contacts" interchangeably to refer to friendship or following (also called subscribing to) relations between users in social networks, and that the focus here is mainly on friendship. A tie means the existence of a direct path between two social network users. It can be argued that since friends of a user have stronger ties with that user (assuming that friendship in online social networks has the same meaning as friendship in the offline world), and consequently a greater similarity, they should participate more in propagating the user's content, and consequently affect its propagation more than non-friends.

According to the literature, similarity is a boosting agent for content propagation, while our previous study (Afrasiabi Rad & Benyoucef, 2012) interestingly showed that strangers (non-friends, and non-followers) affected YouTube content propagation more than friends. Our objective here is to analyze communities (communities are formed by ties between users of a social network, and detected using random walks (Pons & Latapy, 2005)) within the YouTube social network to measure the similarity between members of those communities. For that we compute and analyze similarity metrics within the entire social network, and within its communities. This gives us a comparative tool for investigating similarity values. We also evaluate the ratio of friendship over similarity with the goal of understanding if similar community members are in fact friends.

We focus on interest similarity since it is one of the most effective similarity measures contributing to the propagation of content or influence (Tang, Sun, Wang, & Yang, 2009). Although online social networks differ in their settings and content types, and probably follow different similarity patterns, a look at the work of Mislove et al. (Mislove, Marcon, Gummadi, Druschel, & Bhattacharjee, 2007) leads us to conclude that social networks that fall into the same category based on their privacy settings, user demographics, and applications, display similar information dissemination and similarity patterns. Considering that, we selected YouTube for our analysis as a good representative of online open social networks. We measure interest similarity between YouTube users based on the common topics they share with their friends, followers, and strangers in communities. We measure the similarity of connected and unconnected users in each community, and analyze the ratio of links between similar users versus dissimilar users. This will lead us to answer the question: "do similar users in communities befriend each other, and to what extent?"

Researchers in sociology, mathematics, and physics have proposed different similarity measures, and Social Network Analysis has adopted them to study similarity in social networks. In this paper we evaluate some of these similarity measures in a real social network setting and evaluate them based on the ratio of friendship between similar users.

The rest of the paper is organized as follows. The next section provides an overview of YouTube. Section 3 provides an introduction to the similarity measures used in our study. Section 4 is devoted to the results of our analysis. We continue on with a discussion in Section 5, and conclude the paper in Section 6.

2. Background

Similarity in social networks has been investigated from different angles. McPherson et al. categorized similarity into two categories: status homophily, and value homophily (McPherson, Smith-Lovin, & Cook, 2001). Status homophily can be regarded as structural similarity and value similarity is what we define in our research as interest similarity. According to McPherson et al. value homophily is derived from status homophily, hence it can be concluded that connected parties show similar interests and behavior.

However, Hinds et al. showed that, in a work environment (e.g., corporate social network), value homophily is a stronger indicator of tie formation than status homophily (Hinds, Carley, Krackhardt, & Wholey, 2000), which leads to the conclusion that McPherson et al.'s argument does not hold for every type of social network. However, research on social marketing reports that value homophily is an enabler of word-of-mouth distribution in online social networks (Anderson, 1998; Bernard J. Jansen, Mimi Zhang, Kate Sobel, & Abdur Chowdury, 2010; Hu et al., 2011). The importance of value homophily for online word-of-mouth distribution, hence for tie development, motivates us to investigate the relationship between value homophily and tie creation in online social networks, namely in YouTube as a representative of online social networks.

YouTube: an Open Social Network

YouTube, a subsidiary of Google, is the largest video sharing website containing about 43% of all videos found on the Internet (Flosi, 2010). Since its launch in 2005, the popularity of YouTube has consistently increased, and more web users, from various demographics, registered on this video sharing website to benefit from its contents and features. YouTube is not just an online repository for videos uploaded by users. YouTube also accounts for being a social network since it has a large number of registered users (aka channels) who can upload videos, follow (aka subscribe to) other channels, and be friends with other users (aka channels). Thus, many channels in YouTube have millions of friends and subscribers (YouTube LLC., 2010). YouTube, to fully qualify as a social network, provides facilities that enable communication and interaction between its members. YouTube satisfies this requirement by implementing a broad infrastructure that allows users to communicate with each other in many different ways which resulted in users commenting on nearly 50% of YouTube videos (YouTube LLC., 2010). YouTube's communication infrastructure includes the following features: Private messaging, Commenting on channels, Commenting on videos, Marking a video as favorite (favorite marking), Publishing video descriptions, Liking or disliking a video description or a comment (rating), and Replying to a comment. In reality, users (channels) who subscribe to a channel will receive updates about the channel's activities on their news feed, and whenever they make a comment about or favorite-mark an activity, this act will appear in the news feed of their followers, and, in this way, the activities will propagate in the network.

YouTube also provides APIs that can be used by other web platforms interested in integrating YouTube services. By being integrated with many other web platforms, YouTube videos are not only displayed on a user's profile page, but they can be delivered directly to subscribers, and even the general public (online users) via email, Really Simple Syndication (RSS), and even in connection with other social networking platforms such as LinkedIn and Facebook. Videos can be also searched in search engines such as Google and Bing. These functionalities help YouTube videos to be propagated not only inside YouTube, but also on other platforms, which provides a unique advantage for word-of-mouth distribution, and is actually the reason for us choosing YouTube in our study.

YouTube provides the advantage of allowing two types of ties between channels: friendship, which creates a two-way relationship for channels, and subscription, which allows channels to get updates on any other channel while having a one-way relationship with those channels (Chakrabarti et al., 1999).

Another reason for choosing YouTube in our study is the fact that it allows (as of December 2010) for the existence of groups. By joining different groups, YouTube users could have access to a set of contents of their interests, all gathered in one location. Although Google has decided to revoke access to YouTube groups in December 2010, and has integrated it with Google+, our data, which was collected in 2007, shows a large participation of users in YouTube groups. Hence, we use YouTube group membership as an indicator of the interests of YouTube users. We argue that being members of the same group is indicative of the similarity of interests. In the next section, we explore different similarity measures used to evaluate similarity between users.

3. Similarity Measures and Functions

This section is devoted to a review of popular similarity measures used in social network analysis. According to Lin (Lin, 1998), similarity is a function of commonality and difference, in a way that if two objects are not exactly the same, their similarity depends positively on the amount of their common features, and will have negative relations with their differences.

Many similarity measures have been developed; each tied to an application or requiring a specific domain and design. Therefore, not all similarity measures are suitable to be applied on social networks to compute interest similarity. To measure the similarity of YouTube users, first, we selected a set of similarity measures that can be applied to interest similarity, and then we applied each measure (all of them discussed in this section) as a function of common group memberships of YouTube users. According to Baatarjav et al. , a group in a social network has specific characteristics that match the profiles of most of its members (Baatarjav, Phithakkitnukoon, & Dantu, 2008). Therefore, users who share a set of group memberships should have a similar profile. Note that analyzing similarity based only on group membership may not provide results as accurate as those that can be obtained by semantically analyzing, for instance, the content of users' postings, and considering the demographic information of users.

Jaccard and Dice's Similarity Coefficient

Jaccard and Dice's similarity coefficient measures are specific to measuring set similarity (Dice, 1945; Jaccard, 1901). They were first developed to measure similarities in ecological studies, but their nature of set operations made them applicable for measuring social similarity. They are computed by dividing the intersection of sets over their union. Jaccard and Dice's index can easily be converted to each other and provide monotonic asymmetric results. Therefore, in this paper, we only use Jaccard similarity coefficient for simplicity. Jaccard index is calculated using the following equation:

$$J(U_1, U_2) = \frac{|H_1 \cap H_2|}{|H_1 \cup H_2|} \qquad (1)$$

Where H_1 and H_2 are the group memberships of user U_1 and user U_2, respectively.

Russel and Rao Similarity

Russell and Rao similarity measure (RUSSELL & RAO, 1940) is close to Jaccard's similarity coefficient. Russell and Rao measure the similarity of the common items compared to the whole vector including the attributes, here groups, that are absent from both vectors. In other words, the Russell and Rao similarity measure computes the common group memberships versus the whole set of unique groups in the system, and is calculated by:

$$R(U_1, U_2) = \frac{|H_1 \cap H_2|}{|H|} \qquad (2)$$

Where H represents the total number of group memberships.

Roger and Tanimoto Similarity
Roger and Tanimoto (Rogers & Tanimoto, 1960) devised a measure that is suitable for comparing the similarity of Boolean vectors. Their model gives double weight to disagreements. The Roger and Tanimoto index is calculated by:

$$T(U_1, U_2) = \frac{|H_1 \cap H_2| + |H_1^C \cap H_2^C|}{-3|H_1 \cap H_2| + 2(|H_1| + |H_2|) + |H_1^C \cap H_2^C|} \qquad (3)$$

Where H_1^C represents the groups that do not have user U_1 as their member.

Sokal and Sneath Similarity
Sokal and Sneath similarity measure (Sneath & Sokal, 1973) is comparable to Dice's measure and to Roger and Tanimoto measure. The only difference between Sokal and Sneath and Roger and Tanimoto similarity measures is in the heuristic constant components of the formulas, which produce almost similar results. Sokal and Sneath give double weight to matches instead of differences. Sokal and Sneath, however, founded their model on the Jaccard and Dice similarity measure by extending it to integrate dissimilarity of items into the calculation of similarity. It is calculated by:

$$S(U_1, U_2) = \frac{|H_1 \cap H_2| + |H_1^C \cap H_2^C|}{|H_1 \cap H_2| + |H_1| + |H_2| + 2|H_1^C \cap H_2^C|} \qquad (4)$$

L^1 and L^2 - Norms
With regard to sets, L^1-Norm, and L^2 -Norm (Gradshteyn, Ryzhik, Jeffrey, & Zwillinger, 2000) evaluate similarity to be the overlap between two groups divided by their sizes. L^2 –Norm compared to L^1-Norm decreases the level of effect that the sizes of individual sets have on the similarity measure. L^1 and L^2 –Norms are measured by:

$$L^1(U_1, U_2) = \frac{|H_1 \cap H\_2|}{|H\_1| \cdot |H\_2|} \qquad (5)$$

$$L^2(U_1, U_2) = \frac{|H_1 \cap H\_2|}{\sqrt{|H\_1| \cdot |H\_2|}} \qquad (6)$$

4. Interest Similarity and Ties in YouTube

According to Crandall et al. (Crandall et al., 2008), friends and followers in social networks are either similar to each other at the time the friendship (or follower) tie is made (aka selection process) or they grow in similarity over time after they become friends or followers through social influence. Also, rising similarity between two individuals is an indicator of current, and more specifically future, interactions between them (Crandall et al., 2008; Feld, 1981). Therefore, we argue that current activities of friends and followers of a user, who are presumed to have a certain degree of similarity, can be a predicator of that user's next activity. Hence, friends, also recognized as the most similar people by Crandall et al. (Crandall et al., 2008), should have the greatest effect on content propagation. But the question is: are friends the most similar people in their community? This section attempts to answer this question by analyzing data extracted from YouTube for similarity friendship ratios (the ratio yielding that what percentage of similar users in communities are friends). To do so, we utilize the similarity measures defined in Section 3 of this paper. Note that we cleaned the YouTube dataset to only keep friends in our evaluation and ignored all follower links in order to comply with the findings of Crandall et al. (Crandall et al., 2008) who only consider reciprocated links (here, YouTube friends).

Before we proceed, it is important to comprehend that communities are different from groups, where communities are concepts that are generated based on existing links between social network members, and groups are a feature introduced on social networks to gather users with similar profiles into a single place.

Data Description
Before developing our analysis, the data must be cleaned and made ready for analysis. We have access to a large dataset of over 1.15 million YouTube users and their group memberships along with information about ties between them. This dataset was collected and formerly used in

an analysis by Mislove et al. (Mislove et al., 2007). The dataset covers more than 30 thousand groups and contains over 290 thousands recorded group memberships, so on average, every user in the dataset is a member of roughly four groups. Every user, on average, has more than four reciprocatory and non-reciprocatory ties with other users. The most connected user has over 28 thousand links, while the majority of users only have one link. Figure 1 shows the frequency distribution of ties per user in the YouTube social network.

TABLE 1. YouTube Statistics

Type of Data	Statistics
Users	1,157,827
Groups	30,087
Users That are member of at Least One Group	94,238
Users That are not Members of any Group	1,063,589
Links	4,945,382
Number of Group Memberships	293,360
# of Groups that a user with highest number of membership is subscribed in	1,035
# of memberships for a group that has highest number of memberships	7,591
# of Communities	139,142

The highest number of ties in the network belongs to a user with 28,644 connections while the second most connected user only has 11,239 connections. Interestingly, about 183 thousand users only have one connection, and more than 500 thousand are not connected at all. This shows the level of uneven distribution of inactivity and activity in the YouTube social network. As it is apparent in Figure 1, most users have less than 128 ties. The full statistics of the YouTube dataset used in this study can be found in TABLE 1.

A more detailed look at the statistics shows that about 8% of the users are members of groups, which accounts for about 10 memberships per group. From this point on, our analysis only considers users who are group members, and we simply discard from our analysis the users who did not use YouTube's group feature. The statistical data also illustrates that, on average, users have three common group memberships,

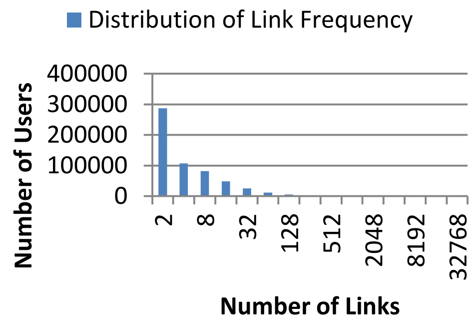


Figure 1. Frequency of ties per user

which shows a great potential for similarity between users.

As planned, we then extracted communities from the YouTube dataset. To do so, we relied on the random walk community detection technique described in (Pons & Latapy, 2005). The Random Walk community detection method discovers communities based on their structural similarity. It first estimates the distance of vertices, as a metric for estimating structural similarity, and assigns it to them as a weight. The next step is applying a hierarchical clustering model in order to identify clusters (communities). The algorithm works at the time complexity of $n^2 \log(n)$, which is suitable for analyzing large graphs. We identified over 139 thousand communities with an average of 11 members per community, the largest community having 73 members.

Analysis of Similarities

As detailed earlier in this paper, we use common group memberships of users in the YouTube social network to measure the similarities between them. We argue that users who are members of the same set of groups are more likely to have similar interests, and that the similarity of interests increases as the number of common group memberships increases.

In order to perform this analysis, we implemented six programs, each of them responsible for performing one similarity measurement operation. The programs performed their analysis on a cleaned database of YouTube users that were previously clustered for communities using our RandomWalk clustering program developed using C++ and the iGraph (www. igraph. sourceforge. net) library.

To measure similarities, we selected six well-defined and generally accepted similarity measures as detailed in Section three of this paper. TABLE 2 describes the result of applying each technique on YouTube social network and its extracted communities.

TABLE 2 shows that for every similarity measure, the similarity of users within the communities is greater than the similarity within the entire social network. Being connected increases similarity, and therefore community members are more similar to each other than the rest of the network.

TABLE 2. Similarity Measures and the Result of Applying Them on the YouTube Social Network and its Communities

Metric	Social Network Average	Average Over Communit-ies
Jaccard	0.14	0.31
Russel and Rao	0.90	0.91
L1	0.12	0.17
L2	0.26	0.34
Sokal and Sneath Similarity	0.50	0.54
Roger and Tanimoto Similarity	0.40	0.47

However, being a member of a community does not necessarily indicate friendship. A community is a collection of users who have transitive connections to each other. Therefore, there is a path between most community members. This also results in a high clustering coefficient for every node in the community. This means that a community is created from the collection of friends, friends of friends and so on. Based on our analysis, it is still not clear how much similarity induces friendship. To be able to answer this question, we selected users who have a more than average similarity with each other in their community, and examined if they are friends or not.

However, being a member of a community does not necessarily indicate friendship. A community is a collection of users who have transitive connections to each other. Therefore, there is a path between most community members. This also results in a high clustering coefficient for every node in the community. This means that a community is created from the collection of friends, friends of friends and so on. Based on our analysis, it is still not clear how much similarity induces friendship. To be able to answer this question, we selected users who have a more than average similarity with each other in their community, and examined if they are friends or not.

The result of our analysis shows that there is not a high correlation between similarity and friendship in communities (see TABLE). In other words, most similar users are not necessarily friends even in small communities within the social network. Note that being in the same community means either a direct friendship or the existence of a short path with many mutual friends between two users. The friendship similarity ratio in small communities of connected people is not large (a range of 11% to maximum 38%). Most similar users in communities are not friends with each other. In our former study (Afrasiabi Rad & Benyoucef, 2012), we observed that content propagation in social network communities is done mostly by non-friends or non-followers. Also, as argued in the literature, content propagation happens where there is a high similarity between the propagator and propagatee. Therefore, it can be deduced that it is possible for indirect friends to be more similar that direct friends. Thus, a comparison of the results presented in TABLE 2 and and TABLE 3 suggests that the higher average of similarity in communities might be the result of high similarity between indirect friends rather than similarity between friends.

TABLE 3. Similarity Friendship Ratios in Social Network and Communities

Metric	Similarity Friendship Ratio in Communities
Jaccard	0.12
Russel and Rao	0.38
L1	0.32
L2	0.11
Sokal and Sneath Similarity	0.12
Roger and Tanimoto Similarity	0.21

5. Discussion

Our analysis shows that every similarity measurement method consistently yielded some degree of similarity between users in communities. Based on the proposition in (Feld, 1981), the higher similarity within communities was expected to be higher than the average similarity in the whole social network. This was confirmed by our results. However, the subsequent analysis that resulted in relatively low friendship similarity ratios in the communities was unexpected. Feld (Feld, 1981) proposes similarity as a determining factor in social ties in offline social networks. Nevertheless, the situation can be different in online social networks. Offline social networks are known to be free of fake friends and spammers which is certainly not the case for online social networks (Manago, Taylor, & Greenfield, 2012). The problem starts to grow when we realize that fake friends have on average six times more friends than legitimate users (i.e., users whose friends are real) (Manago et al., 2012). Therefore, unless we have a mechanism to separate fake friends from real friends, the results cannot show the true ratio. Nonetheless, the friendship similarity ratio is so low that the general finding of low similarity between friends stands even if fake friends are removed from the network. The only difference would be a slight increase in the ratio.

Based on the research done by Feld (Feld, 1981), it is expected that, in offline social networks, similar people be friends with each other. Our study on YouTube found that this is not necessarily the case for online social networks. However, considering Feld's study, we expect that friends should have higher similarity. Therefore, similarity measures that result in a higher ratio between friendship and similarity provide more accurate results in the case of online social network.

By looking at the results presented in TABLE , the similarity measures that resulted in higher values of friendship similarity ratios in communities are Russel and Rao and L1 similarities. We have a second category including Jaccard, L2, and Sokal and Sneath Similarity, with relatively similar results. Comparing these results with the values presented in TABLE 2, we see that even though the similarity values resulting from different techniques vary, the techniques can be categorized into two major categories with regards to their approximate accuracy. A conclusion about which category provides better results will depend on more research to be conducted on the correlation between friendship and similarity in online social networks. In which case, a higher correlation will play in favor of the first category of measurement techniques, and a lower correlation will favor the second category.

6. Conclusion

In this paper we analyzed the YouTube social network with regards to the ties that exist between users and their common group memberships (which we used as an indicator of similarity of interests), to assess the relation between friendship and the similarity of interest inside communities of users within a social network. We found that the similarity between users increases if they are friends, but this increase does not define similarity as a determining factor in friendship.

Considering that, and also the fact that content propagation in online social network communities is done mostly by non-friends, and knowing that similarity is a driver for content propagation, we can conclude that, within communities, indirect friends are more similar to each other than direct friends (as they participate more in content propagation). The second possibility is that the YouTube communities are formed from users that have little similarity whether friends or non-friends. The deterministic conclusion on the findings discussed above needs more exploration on the similarities between indirect friends, which is one the paths for our future study.

Furthermore, we examined several similarity measures to find the most suitable ones for processing online social network data. We found that similarity measures can be categorized into two categories based on their accuracy, which is measured by the friendship ratio. The results yielded by the Russel and Rao as well as L1 similarity measures led to higher friendship similarity ratio, and Jaccard, L2, and Sokal and Sneath Similarity fell in the second category. More research is needed to determine which category provides better results for online social networks.

Our analysis can be developed further to extract better facts from a social network like YouTube. One of the limitations of this research is the lack of comprehensive data on the YouTube network. We only used a sample of YouTube, where users are group members, and we ignored users who are not members of a group. This resulted in a large YouTube user base. Therefore, a higher group membership rate would have improved the results.

In our future work, we plan to investigate the validity of our findings on different types of social networks, such as photo sharing (Flickr), friendship (Orkut), professional (LiveJournal), and so on. Furthermore, we will try to detect fake friendships and remove them from our analysis to obtain more accurate results.

7. References

Afrasiabi Rad, A., & Benyoucef, M. (2012). Measuring Propagation in Online Social Networks: The case of YouTube. *Journal of Information Systems Applied Research*, *5*(1), 26.

Anderson, E. W. (1998). Customer Satisfaction and Word of Mouth. *Journal of Service Research*, *1*(1), 5–17. doi:10.1177/109467059800100102

Asur, S., & Huberman, B. A. (2010). Predicting the Future with Social Media. *SSRN eLibrary*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1579522

Baatarjav, E.-A., Phithakkitnukoon, S., & Dantu, R. (2008). Group Recommendation System for Facebook. In R. Meersman, Z. Tari, & P. Herrero (Eds.), *On the Move to Meaningful Internet Systems: OTM 2008 Workshops*, Lecture Notes in Computer Science (pp. 211–219). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-88875-8_41

Bearden, W. O., Calcich, S. E., Netemeyer, R., & Teel, J. E. (1986). An Exploratory Investigation Of Consumer Innovativeness And Interpersonal Influences. *Advances in Consumer Research*, *13*(1), 77–82.

Bernard J. Jansen, Mimi Zhang, Kate Sobel, & Abdur Chowdury. (2010, March 6). Twitter Power: Tweets as Electronic Word of Mouth. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.155.3321

Chakrabarti, S., Dom, B. E., Kumar, S. R., Raghavan, P., Rajagopalan, S., Tomkins, A., Gibson, D., et al. (1999). Mining the Web's link structure. *Computer*, *32*(8), 60 –67. doi:10.1109/2.781636

Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., & Suri, S. (2008). Feedback effects between similarity and social influence in online communities. *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '08 (pp. 160–168). New York, NY, USA: ACM. doi:10.1145/1401890.1401914

Dice, L. R. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology*, *26*(3), 297–302. doi:10.2307/1932409

Domingos, P., & Richardson, M. (2001). Mining the network value of customers. *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '01 (pp. 57–66). New York, NY, USA: ACM. doi:10.1145/502512.502525

Duan, W., Gu, B., & Whinston, A. B. (2008). Do online reviews matter? - An empirical investigation of panel data. *Decis. Support Syst.*, *45*(4), 1007–1016.

Evans, D. C. (2009, January 15). *Beyond Influencers: Social Network Properties and Viral Marketing*. Psychster Inc. Retrieved from http://www.slideshare.net/idealistdave/beyond-influencers-social-network-properties-and-viral-marketing-presentation

Facebook Inc. (2011). Facebook Statistics. Retrieved July 9, 2011, from http://www.facebook.com/press/info.php?statistics

Feld, S. (1981). The Focused Organization of Social Ties. *American Journal of Sociology*, *86*(5), 1015–1036.

Flosi, S. (2010, June). comScore Releases May 2010 U.S. Online Video Rankings - comScore, Inc. *ComScore Inc. - Press release*. Retrieved May 4, 2012, from http://www.comscore.com/Press_Events/Pre ss_Releases/2010/6/comScore_Releases_May_2010_U.S._Online_Video_Rankings

Gradshteyn, I. S., Ryzhik, I. M., Jeffrey, A., & Zwillinger, D. (2000). *Table of Integrals, Series, and Products, Sixth Edition* (6th ed.). Academic Press.

Hinds, P. J., Carley, K. M., Krackhardt, D., & Wholey, D. (2000). Choosing Work Group Members: Balancing Similarity, Competence, and Familiarity. *Organizational Behavior and Human Decision Processes*, *81*(2), 226–251. doi:10.1006/obhd.1999.2875

Hu, N., Tian, G., Liu, L., Liang, B., & Gao, Y. (2011). Do Links Matter? An Investigation of the Impact of Consumer Feedback, Recommendation Networks, and Price Bundling on Sales. *Engineering Management, IEEE Transactions on*, *PP*(99), 1–12. doi:10.1109/TEM.2010.2064318

Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin del la Société Vaudoise des Sciences Naturelles*, *37*, 547–579.

Judea, P. (1986). Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, *29*(3), 241–288. doi:10.1016/0004-3702(86)90072-X

Kempe, D., Kleinberg, J., & Tardos, É. (2005). Influential Nodes in a Diffusion Model for Social Networks. *Automata, Languages and Programming* (pp. 1127–1138). Retrieved from http://dx.doi.org/10.1007/11523468_91

Kim, Y. A., & Srivastava, J. (2007). Impact of social influence in e-commerce decision making. *Proceedings of the ninth international conference on Electronic commerce* (pp. 293–302). Minneapolis, MN, USA: ACM. doi:10.1145/1282100.1282157

Leskovec, J., Adamic, L. A., & Huberman, B. A. (2007). The dynamics of viral marketing. *ACM Trans. Web*, *1*(1), 5+. doi:10.1145/1232722.1232727

Lin, D. (1998). An information-theoretic definition of similarity. *Proc. 15th International Conf. on Machine Learning* (pp. 296–304). Morgan Kaufmann, San Francisco, CA. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.55.1832

Manago, A. M., Taylor, T., & Greenfield, P. M. (2012). Me and my 400 friends: the anatomy of college students' Facebook networks, their communication patterns, and well-being. *Developmental Psychology*, *48*(2), 369–380. doi:10.1037/a0026338

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, *27*, 415–444. doi:10.2307/2678628

Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., & Bhattacharjee, B. (2007). Measurement and analysis of online social networks. *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement* (pp. 29–42). San Diego, California, USA: ACM. doi:10.1145/1298306.1298311

Pons, P., & Latapy, M. (2005). Computing Communities in Large Networks Using Random Walks. In pInar Yolum, T. Güngör, F. Gürgen, & C. Özturan (Eds.), *Computer and Information Sciences - ISCIS 2005*, Lecture Notes in Computer Science (pp. 284–293). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/11569596_31

Rad, A., Amir, & Benyoucef, M. (2011). A Model for Understanding Social Commerce. *Journal of Information Systems Applied Research*, *4*(2), 63.

Rogers, D. J., & Tanimoto, T. T. (1960). A Computer Program for Classifying Plants. *Science*, *132*(3434), 1115–1118. doi:10.1126/science.132.3434.1115

RUSSELL, P. F., & RAO, T. R. (1940). On Habitat and Association of Species of Anopheline Larvae in South-Eastern Madras. *Journal of the Malaria Institute of India*, *3*(1), 153–178 .

Sneath, P. H. A., & Sokal, R. R. (1973). Numerical taxonomy. The principles and practice of numerical classification. Retrieved from http://www.cabdirect.org/abstracts/19730310919.html

Stephen, A. T., & Toubia, O. (2009). Deriving Value from Social Commerce Networks. *SSRN eLibrary*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1150995

Tang, J., Sun, J., Wang, C., & Yang, Z. (2009). Social influence analysis in large-scale networks. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '09 (pp. 807–816). New York, NY, USA: ACM. doi:10.1145/1557019.1557108

Trusov, M., Bodapati, A. V., & Bucklin, R. E. (2010). Determining Influential Users in Internet Social Networks. *Journal of Marketing Research*, *XLVII*(August), 643–658. doi:10.2139/ssrn.1479689

Twitter Inc. (2011, March 14). Twitter Blog: #numbers. *Twitter Blog*. Blog. Retrieved March 8, 2012, from http://blog.twitter.com/2011/03/numbers.html

Van den Bulte, C., & Joshi, Y. V. (2007). New Product Diffusion with Influentials and Imitators. *MARKETING SCIENCE*, *26*(3), 400–421. doi:10.1287/mksc.1060.0224

YouTube LLC. (2010). YouTube - Press Statistics. Retrieved July 9, 2011, from http://www.youtube.com/t/press_statistics

Looking Backwards to Look Ahead: Lessons from Barcode Adoption for RFID Adoption and Implementation

Aditya Sharma
asharma@nccu.edu
Computer Information Systems
School of Business
North Carolina Central University
Durham, NC 27707

Dominic Thomas
dthomas@suffolk.edu
Information Systems & Operations Management
Sawyer Business School
Suffolk University, Boston, MA 02108

Abstract

This paper compares the adoption patterns of two automatic identification technologies i.e. Bar codes and RFID (Radio Frequency Identification). The paper juxtaposes the historical events that were significant in the adoption of Bar codes with the contemporary events that are taking place in the RFID space. Based upon the review of bar coding literature and data collected from semi-structured interviews, the paper identifies critical themes and eight key enablers underlying the adoption of bar codes and suggests how understanding of those themes and enablers can inform the adoption and implementation of RFID and similar emerging technologies.

Keywords: Adoption, Automatic Id, RFID, Barcodes, Platform Innovation, Emerging Technology

1. INTRODUCTION

Automatic Identification is the process of identifying and tracking objects through the use of technology devices such as magnetic readers, bar codes and radio frequency. While keyless data entry devices have existed since 1800's when they were used as reading aids for the blind, the invention of electronic digital computers led to the search for better methods of data entry (LaMoreaux, 1998).

With increase in the logistics and inventory costs for supermarkets in the mid-1900's there was a growing need to find an efficient means for automatic identification of products without manual inspection. Two graduate students at Drexel institute solved the problem by relating it to Morse code in which messages sent as dots and dashes were read automatically leading to the birth of the Bar code.

RFID (Radio Frequency Identification) is a means of automatic identification of objects using radio signals. While it has been around since the

1940's its commercial application is relatively recent.

In this paper we juxtapose the historical events in the adoption of bar codes and compare them with RFID adoption patterns. We suggest eight key enablers that were critical in Bar Code adoption and that also inform on the adoption and implementation of RFID technologies.

Bar Code

Bar codes were invented in 1949 and by 1952 Norman Joseph Woodland and Bernard Silver were issued the first patent for a bar code type product. The first commercial use of the Bar code was in 1967 when RCA installed them on the first scanning systems at a Kroger Store in Cincinnati. It soon became apparent that an industry standard was needed so that different equipment manufacturers, food producers and dealers could readily adopt it. In 1969 a consortium of food distribution trade associations called the Uniform Code Council (UCC) began to develop a standardized barcode for consumer items called the Universal Product Code or the UPC. In 1973, an Ad Hoc committee composed of grocery industry executives chose the 11-digit, linear bar code that is now commonly referred to as UPC (Haberman, 2001). The initial UPC was a linear one-dimensional bar code, which contained manufacturer and brand information but no uniquely identifying data. In 1974, there was agreement in the UCC on adopting a common standard for the UPC. Thus began the new era of automatic identification of consumer products. While most barcodes are still one-dimensional like the original ones, two-dimensional bar codes, which can carry more data in a smaller area, are commonly used in shipping markets and transit companies such as UPS and FedEx (Ucc Website).

Bar codes suffer from several limitations. Objects must be physically manipulated to align with scanners to get a line of sight. Barcodes are exposed to vagaries of the environment and with natural wear and tear become inefficient. This is quite evident when many times checkers face difficulties in scanning an item. Bar codes require sequential processing of data and need to be brought in line of sight of the scanner, one item at a time. Also bar codes carry limited data, which is static in nature; hence the identification is usually at the product level unless special efforts are made to identify the item. Radio Frequency Identification or RFID has the potential to alleviate the problems presented by barcodes.

RFID Technology

The United States Air Force developed RFID technology in the 1940s to differentiate between friendly and enemy aircraft in World War II. Though patented in 1973, it has only become commercially and technologically viable for commercial applications in recent years. As compared to barcodes RFID has the potential to provide improved data collection and handling through more granular data, geospatial/physical alignment independence, parallel processing of multiple scans simultaneously, and internal placement in objects.

Basic identification data is carried in transponders known as tags, read by transceivers that decode and transmit data to attached computers for processing. There it can be associated with database information such as product, business processes and organization data. The data in a tag (also referred to as tag id) can identify the object associated with it in terms of its manufacturer, brand, model and unique serial number for the object. Thus data are granular to the specific product level. The tag consists of a small microchip attached to an antenna and communicates via radio frequency with a transceiver or tag reader. A tag has geospatial/physical alignment independence in that it may be read without any line of sight. Tags can be read at a rate of several hundred reads per second (essentially simultaneous) and from a distance of several meters. The tag can be attached to the outside or the inside of a product that is made of non-conducting material, without read problems or wear and tear. RFID tags have a unique ability to be active (battery power source) and can be combined with other technologies to capture contextual information such as temperature variations to create a history of the object through its life cycle.

Up until now, RFID has been too expensive and too limited in adoption levels to be practical for many commercial applications. With recent reduction in tag and RFID systems costs, RFID can solve many of the problems associated with barcodes. Unlike barcodes RFID does not require a "line of sight" to track products and no manual intervention is needed. Radio waves travel through most non-metallic materials except liquids, so they can be embedded in packaging or encased in protective plastic for weatherproofing and greater durability.

Additionally, tags have microchips that can store a unique serial number for every product manufactured around the world and can also be updated.

Business Impacts and Benefits from RFID

The power of RFID lies in its ability to capture or acquire more data, automatically without manual intervention, in almost real time. The data can be the unique identity of each item in its location and could potentially help in tracking the item in real time and creating rich profiles, which could be the history of the object from its time of creation to its eventual destruction. The physical object is no longer an abstraction of reality but tied to reality itself. The data is available at the item-level and multiple items can be scanned simultaneously using radio waves. This empowers businesses by allowing them to create automated inventory control systems, enabling real time inventory management, and therefore making their supply chains more efficient. Database updates could occur in real time, resulting in more dynamic systems. This is analogous to having a live video versus a snapshot of the process in time.

The potential benefits from RFID for consumer product applications relate to ease of use. Manufacturers, transporters and retailers scan millions of bar codes every day; however each may use their own formats, and usually the bar codes are scanned only at a single point, such as checkout, due to the processing burden of arranging manual orientation and line of sight. By integrating RFID at each level in a supply chain, every party involved in the lifespan of a product can potentially scan every product within a scanner-enabled supply chain location at any time. This includes not only manufacturers or retailers but also regulatory bodies such as the FDA, end consumers and even waste disposal and recycling organizations. RFID has the potential to lower costs of inventory management, supply chain management and retail checkouts as no individual worker need be present during a scanning.

If used in this manner, RFID technology will provide “real time” information in tracking products and opportunities for creating rich product life-cycle profiles. These could be used to increase theft prevention, inventory management accuracy and quality control. Besides these three apparent direct benefits, RFID deployment can result in many indirect benefits such as better business customer management, enhanced partner collaboration, and more efficient business processes resulting from process mapping and through gaining strategic insight into product-level life-cycles.

Over the last decade, RFID has been implemented to improve goods tracking throughout supply chains (SC), access control for security, livestock management, waste management tracking, inventory control, and transportation fleet management. As RFID use grows in its trajectory of becoming a commonly adopted technology, firms have begun thinking up new ways of leveraging RFID’s technological capabilities. One forefront in these innovations will be making active RFID tags, which can store and provide rich status information from sensors on tagged items.

Leading retailers such as Wal-Mart and Target and manufacturers such as Proctor and Gamble and Gillette have endorsed the technology and are pilot-testing its use for full-scale retail implementation.

2. BAR CODE AND RFID ADOPTION TIMELINES

While the commercial use of bar codes began in 1974, the adoption of bar codes did not pick up until the early 1980’s when mass retailers K-Mart pushed for its adoption. It took nearly 20 years for full-scale adoption of bar codes. Adoption of RFID is likely to follow a similar pattern but with a shorter time cycle. This reduction in time is likely due to advances in information technology and quicker responses to environmental forces. Even though RFID has been around for many years, its commercial application has been relatively recent and has picked up only in the later part of 1990’s and early 2000’s. Table 1 and Table 2 in the Appendix present the timeline of critical events for Bar Code and RFID adoption.

3. DATA COLLECTION & ANALYSIS

We wished to explore the key enablers in the adoption of RFID by organizations and understand what factors were contributors or deterrents and may impact their decision to adopt and integrate RFID internally. We were curious not only about the decision to adopt but also whether the organizations intended to integrate data generated by RFID with internal systems and processes. In such a case,

interpretive research focusing on exploring the unknown phenomenon best serves to initiate a valid and accurate line of inquiry (Krippendorff, 1980) precisely our underlying research goal. To accomplish the above-mentioned goals and to develop a better understanding of the adoption process, we conducted in-depth, semi-structured interviews using a convenience sample.
The interviewees were executives and RFID program managers and supply chain managers across 10 organizations (12 interviews) involved in RFID initiatives at some level. We sampled from three perspectives in order to triangulate and, thereby, strengthen our understanding of RFID adoptions. These perspectives were the adopter perspective (7 firms and 8 interviews in three industries: manufacturing, retailing, and logistics), the implementer perspective (1 top IT consulting firms and 2 interviews), and the vendor perspective (2 firms and 2 interviews). Table 3 in the Appendix describes the profiles of organizations interviewed and their decision status on RFID Adoption and Integration.

The interviews were conducted over a period of three months (May-July, 2005) and were either face to face or over the phone, lasting between one and two hours. The questions for the interviews were a mix of open-ended questions and closed questions to allow both the flexibility of exploring new contexts but also to help maintain focus on some of the previously identified relevant themes from bar code adoption and prior literature. These themes emerged from the data and were later developed conceptually, because of what we found from practice.

The interviews were recorded and later transcribed. The author coded the interview data in an effort to extract key ideas underlying the decision to adopt RFID for managers evaluating emerging technologies such as RFID. This coding process involved the first author identifying patterns and underlying themes that emerged from quotations in the raw text, excerpting them and bringing them to the other author for joint discussion and refinement over a period of 7 months and more than 20 hours of discussion.

4. KEY RFID ADOPTION ENABLERS

In executing this study comparing the adoption vector of barcodes 30 years ago with RFID today in the commercial arena we have been able to extrapolate eight key enablers and evaluate their current status in RFID settings. In addition to the literature review to collect information for comparison, we conducted interviews with managers in charge of RFID research and implementation efforts at 10 firms in industries ranging from logistics and manufacturing to marketing and retail to find out their current outlook on each of the eight enablers. We present these findings in the following sub-sections as a guide for those involved in RFID projects or otherwise interested in successful RFID implementation and adoption in commercial applications. Table 4 in the Appendix summarizes these findings.

Establishing the Standards
Development of standards is critical in the adoption of any new emerging technology. Prior research on standardization has suggested that standardization emerges as a result of an inter-firm cooperation strategy. This theme of literature has examined the incentives to technological compatibility (Besen & Farrell, 1994); collective nature of organizational action in the emergence of standards (Vab De Ven & Garud, 1989) and the governance of collaborative standardization (Antonelli, 1994). In the case of bar codes for an automated checkout system to work, supermarkets and packaged goods companies had to agree on one standard to translate lines into numbers representing the same product-model consistently to avoid confusion. The grocery industry realized this challenge early on and created an Ad Hoc committee with representatives from different groups (i.e. manufacturers, distributors and retailers) in 1970. The Ad Hoc committee worked towards accomplishing the goal of a common standard. Finally, in 1973 through the efforts of the Ad Hoc committee representatives of supermarkets and their counterparts from consumer-goods companies agreed upon the Universal Product Code (UPC) to handle the issue of data compatibility.

In the case of RFID standards or rather the lack thereof, companies appear to be adopting a wait and watch approach thus further delaying adoption. As interviewee from organization A which is a Home construction retailer stated, *"We find benefits but RFID is not on our priority list and we don't think we are ready as we don't have the infrastructure and expertise to process huge amount of data that would be generated by it and make sense out of it. Lack of standards* *and cost of tags and readers is prohibitive."* Besides hardware, software, and middleware standards, another important issue that needs to be dealt with is the adoption of legal standards and intellectual property rights incorporating potential points of contention such as who owns the tags, can they be deactivated, and the management of information on the tags. All these legal-property rights aspects may delay adoption further. Many business and technology experts expect that resolving these standards and legal-property rights issues may help in accelerating RFID adoption. The proactive role of standard making body EPC global and the movement towards the GEN2 standard is likely to promote more widespread adoption of RFID.

Solving the Chicken & Egg Dilemma: Network Effects, Critical Mass & Economies of Scale

The adoption of bar codes posed the classic chicken and the egg problem. Why would manufacturers put bar code on their product if there were no retailers to scan it? And why would retailers invest in scanning equipment unless a significant amount of their product was coded (Brown, 1997). This scenario is similar to the adoption of any technology that exhibits network effects i.e. the greater the number of adopters of the technology, the more beneficial it becomes for its users. Prior research has indicated that in technologies exhibiting network externalities adoption may be driven through sponsorship and support (Katz & Shapiro, 1986; Riggins, Kriebel & Mukhopadhyay, 1994)]. Adoption of an emerging technology needs to attain a critical mass before the technology can really take off (Markus, 1987).

The Ad Hoc committee recognized the effect of network externalities and the need to attain critical mass. It was their leadership efforts in convincing the groups involved that led to the diffusion of barcodes. The adoption of barcodes by 1350 manufacturers led to almost a ten-fold increase in the probability of adoption of scanners by retailers. Similarly the adoption of scanning by 360 retailers led to a significant increase in the probability of adoption by manufacturers (Haberman, 2001). This scenario is likely to play out also in the case of RFID adoption with similar network effects and with benefits to gain for all from full-scale supply chain integration. As interviewee from consulting firm B mentioned, *"My definition of adoption is a continuum. The continuum has to do with the amount of integration you are putting into your business product. So slap and ship with absolutely no integration what so ever, they are either return lifted or on your return data all the way to a fully integrated solution where you are tracking tags through your supply chain individually."* Also, greater demands for the tags would result in economies of scale in its production and further reduction in tag costs. This is likely to have a cascading effect as reduced tag costs are likely to further drive adoption. The standard making body EPC global (enabler 1) may need to take lead to help cross the critical mass barrier. Another component of affecting this enabler is drive from dominant market players (enabler 3).

Dominant Market Players Driving (Mandates)

Initial bar-code adoption was very limited. In March 1976, Business Week published an article titled 'The Supermarket Scanner That Failed' . (Haberman, 2001). It was widely believed even though incorrectly that the experts had predicted 5,000 stores with scanners by 1975 instead of the 100 that were actually there. This misperception was caused because the experts had estimated that the savings from scanning would justify the investments if there were 5000 stores by 1975. It was only in the early-mid 1980's that bar codes really took off. According to Stephen Brown (1997), "What really turned the corner was not the grocery industry, but the mass merchandisers. When the mass merchandisers, most notably Kmart, decided to adopt the system, that built a momentum that never stopped."

The prophecies of doom and gloom are not new to RFID. Many consider the technology over-hyped. This was no different at the time of the bar code. Wal-Mart provided market leadership in the adoption of UPC, or universal product codes, and is exhibiting the same leadership in the adoption of RFID technology by mandating its adoption among its top suppliers. *"What I see happening now is that Wal-Mart is clearly the biggest driver in RFID technology in the business area. There are a couple other drivers in let's say Pharma. There is chain of custody and issues around counterfeiting and safety for consumers that are some very important issues for the pharmaceutical groups. From a general retail and consumer package goods manufactured perspective, Wal-Mart is the 800-pound gorilla and Wal-Mart is driving"* according to the consulting manager from organization I. Partner mandates are an important driving force for adoption, but mandates alone were not enough to drive barcode adoption and may not be enough to drive suppliers towards full scale integration of RFID without a clear innovation focus (enabler 4).

Focusing on the Innovation Opportunity: Business Case & ROI

It has been mentioned in many current business reports that while the retailers might benefit from the adoption of RFID the suppliers do not find a compelling business case to adopt. A similar scenario played out 35 years ago in the case of bar codes. The Ad Hoc committee provided leadership to identify direct and indirect short-term and long-term benefits for manufacturers, retailers and distributors. Economic benefits and ROI were not realized in the first few years until wide acceptance and adoption of the UPC code (Haberman, 2001). Also, while the barcodes were initially intended to automate checkouts and be beneficial for the retailers some of the major benefits such as operational efficiencies and information management along the supply chain became apparent later.

At present some suppliers view RFID as an opportunity and integrate the technology with their internal processes in order to re-engineer them and make them more efficient. According to the manager of end-user organization G, *"we recently created as a company, which is called the Innovation Experience. It is like if you go to trade shows, they always have a lot of booths and different things, different technology. As a company, we recognize how important it is to show and also allow different businesses to recognize the new technology. We then invite our business to see how they can relate it to their existing processes"*. Some other suppliers are using RFID to be more attractive and appealing to their customers. However many suppliers, those that are only tagging and shipping to meet the mandates are viewing RFID as the cost of doing business rather than a strategic resource. These suppliers are unlikely to see any immediate benefits until widespread adoption takes place. At this intermediate level for enabler three, we believe the persuasive role of standard making bodies (enabler 1) and partnership collaboration (enabler 5) are critical to move to the next level of maturation.

Collaborating among Partners

Barcode adoption required leadership and direction from the Ad Hoc committee which was an inter-industry committee representing the grocery industry. The committee was formed of chief executives with five representatives from grocery manufacturers and five from distributor associations, which included two chains and a wholesaler (Brown, 1997). This grouping enabled a collaborative approach towards solving the problems faced in the adoption of the technology. The interests of manufacturers, retailers and distributors were all given due importance. Also the people representing the committees had decision making power.

When RFID information is used across supply chains with inventory management systems it becomes an interorganizational tool with greater potential benefits derived from increased partner participation and commitment and subsequent refinements of organizational processes. The ability to acquire and communicate unique and relevant information about tagged items/entities at any given place and time almost instantaneously, gives RFID technology the potential to reduce costs, increase operational efficiency and improve performance. Interview data from organizations indicates that partner collaboration is already happening. The consulting manager from organization I, talking about a dominant retail partner and its suppliers said, *"What they are trying to do is to take all of this data and provide it back to their suppliers and say, "You manage your product better within our stores and supply chain for us. You need to tell us when more efficient ways are to be ordered. You need to manage your promotion more effectively.""* Associated information could be used in many different ways to understand and improve processes and significantly enhance competitiveness. Hence it is very likely that due to RFID's trans-supply-chain benefits, adoption will be more effective when done jointly through collaborative arrangements (Yang & Jarvenpaa, 2005) rather than individually by isolated partners.

Partner relationships in a dyad or their memberships to professional and standard making bodies may play an important part in driving adoption. At this point these relationships appear to be nascent for RFID. Wal-Mart and other leading proponents of the technology should work closely and collaboratively with their partners and also use standard making organizations such as EPC global (enabler 1) as a platform to address the concerns of their suppliers.

Developing Supporting Technologies

In the case of bar codes, despite being invented in 1949, barcode technology did not take off for over two decades because lasers and computers were very expensive and were required to effectively derive significant benefits from bar code adoption. This exemplifies the need for supporting technologies to co-evolve in order to support infrastructure type technologies such as bar codes and RFID. Large benefits are expected from the integration of RFID with other applications and systems. Thus, RFID will require the development of new hardware, software and middleware for full-scale integration of the technology with existing systems. As mentioned by manager from organization G which is in the paper industry, *"Ideally, you want to have more re-points in between and have the third party (3pl) the logistics company. The D.C. also has a responsibility on it. So, you can see points and that is how the tracking is supposed to work but until the whole infrastructure is ready and until the EPC-Network is ready; it is very spotty."*

At present it is unclear what all of these additional technologies will be, but some of them may be better batteries at lower cost and sensors and memory for active tags. In the long run, once RFID matures, supporting technologies will flourish as adoption will accelerate (enabler 2). At present this is an enabler that can be refined by building the business cases (enabler 4) among collaborating partners (enabler 5) and with standards bodies (enabler 1).

Addressing Consumer Concerns

During early bar code adoption, consumers were not ready to accept products without price stickers, as they did not trust that retailers would not change prices behind their backs. Issues of trust forced several state legislatures to pass laws mandating price labels thus forcing retailers to continue putting price stickers on products. Similarly in RFID adoption, concerns about consumer privacy issues are rampant. Some consumers fear that all of what they purchase can be scanned easily by someone outside their house. Thus, it is important to provide information and educate consumers on what RFID technology can or cannot do and demonstrate that some of their concerns are unfounded. The interview data suggests that these concerns are more perception than reality but still need to be addressed. According to RFID program manager from Antenna and Label maker organization D, *"There are always folks who have those concerns. And some are legitimate to what I would call infrequent examples of filtering personal data. But really nothing to date that has been significant with RFID. Frankly in terms of personal security you take more risk in handing clerk your credit card than putting an RFID tag around it. Now it's because of the press around those issues a lot of the venders, technology providers, and standards organizations are heavily investing time in security systems."* As the tags become more pervasive, due to the network effect RFID consumer concerns should mimic the pattern experienced by barcodes and decrease quickly, but the initial hurdle remains fairly high at this point (enabler 2).

Acknowledging Likely Unforeseen Impacts: Preparing for Radical Innovation

A new technology provides the opportunity to innovate. For many it could be an opportunity to leap ahead of their competition and many times its "real" benefits might be too "radical" to see upfront. Barcodes were initially seen as a means for automating supermarket checkouts but their "real" value was information creation. This idea was highlighted in an article by Fortune magazine in 2004. Following is a quote from the magazine article, "As sometimes happens with seemingly minor technological changes, bar codes have had a huge and unexpected impact. Previously, cash registers had been mere repositories of money; post-UPC, they became data conduits. Each time a product is sold, a record of the item is now preserved. This altered the balance of power between retailers and manufacturers. Once, manufacturers controlled data about product sales via warehouse inventories. They knew more about the products that were selling than the retailers. But, with UPC barcode adoption, stores now had data too—and both sides would learn to mine that information." (Varchaver 2004)

In the context of RFID technology, it is an infrastructure type technology (Curtin, Kauffman & Riggins, 2007) or a platform innovation that mandates future follow-on investments and significant changes in the routines and practices of organizations to realize benefits. *"Smaller organizations see RFID as an opportunity to make two leaps at once and hence displace some of the existing organizations. For us, in terms of retail checkout at this point it is not a major change, as it does not fundamentally change the business process. But going into the future, when there is item level tagging, and* *automated checkouts, it may be a paradigm shift because it eliminates the basis of our business. We may have to kiss our scanning and retail business goodbye"* according to the interviewee from organization J. As is also true with most infrastructure type innovations such as electricity, it has a much broader impact potential where significant strategic benefits would come more from how the technology is applied. This would mean making significant changes for the organizations and acquiring new knowledge about the innovation and its application in the business settings.

In RFID adoption one must modify business processes to leverage the benefits over barcode and other automated identification technologies. These processes will include inter-organizational processes as well if organizations want to enable real-time insight at a granular level. In this sense adoption of RFID may be characterized as disruptive or radical as it brings about changes in structure and functioning of the organizational entity and its inter-organizational systems. RFID program manager from a logistics and transportation company C mentioned the dilemma that his organization is facing. According to him, *"RFID would require altering our existing optical scanners infrastructure and processes currently in place. A lot of learning, major changes in infrastructure may be required. This would be disruptive for the organization."*

To realize these inter-organizational benefits from RFID adoption, synergies need to be built between organizations collaborating at some level (enabler 5), and they need to be prepared for the unforeseen process impacts RFID implementation may require.

5. CONCLUSION

RFID presently exists in an early stage of maturation as far as commercial applications and adoption are concerned. We identified eight enablers for RFID adoption extrapolated from the successful maturation and adoption of barcode technologies and supported by data from interviews. We present these enablers as guides for those interested in implementing and innovating using RFID technology.

6. REFERENCES

Antonelli, C. (1994). "Localized technological change and the evolution of standards as economic institutions" *Information Economics and Policy,* 6, 195-216.

Besen, Stanley M. and Joseph Farrell. (1994). "Choosing How to Compete: Strategies and Tactics in Standardization, *Journal of Economic Perspectives*, vol. 8, no. 2, pp. 117-131.

Brown, Stephen A. (1997). "Revolution at the Checkout Counter", *Harvard University Press*, Cambridge, MA

Curtin, J., Kauffman, R.J. & Riggins, F.J. (2007). "Making the 'MOST' out of RFID Technology: A research agenda for the study of the adoption, usage and impact of RFID", *Information Technology and Management*, Vol: 8 (2)

Haberman Alan L. (2001). "Twenty Five Years Behind Bars", *Harvard University Press*, Cambridge, MA

Katz M. and Shapiro C. (1986). "Technology adoption in the presence of network externalities," *Journal of Political Economy,* vol. 94, no. 4, pp. 822-841

Krippendorff (1980). Content Analysis: An Intro. to Methodology, Sage Publication, Newbury Park, CA.

LaMoreaux Robert D. (1998). "Bar Codes and other Automatic Identification Systems", Pira International, Surrey, UK

Markus Lynne M. (1987). "Toward a "Critical Mass Theory of Interactive Media" Communication Research, Vol. 14, No. 5, 491-511

Riggins F., Kriebel C. and Mukhopadhyay T. (1994). "The Growth of Interorganizational Systems in the Presence of Network Externalities" Management Science, 40(8)

Ucc Website: http://www.uc-council.org

Van de Ven, A. H., and Garud, R. (1989). "The co-evolution of technical and institutional events in the development of an innovation", in J. Baum and J. Singh (eds), Evolutionary Dynamics of Organizations, pp425-443

Varchaver N. (2004). Fortune Magazine Website:

http://money.cnn.com/magazines/fortune/fortune_archive/2004/05/31/370719/

Yang, G. and Jarvenpaa, S.L. (2005). "Trust and Radio Frequency Identification (RFID) Adoption within an Alliance." In R. Sprague (Ed.), Proceedings of the 38th Hawaii International Conference on Systems Science Big Island, HI

Appendix

Table 1. Bar Code Adoption Timeline	
Year	**Key Events**
1949	Invention of Barcodes
1952	Patent issued
1966	First commercial Application of Bar Codes
1969	Grocery manufacturers and retail associations perceive a need to develop a standardized bar code or UPC code
1970	Grocery Industry Ad Hoc committee formed for developing standards
1973	UCC adopts common standards on UPC
1974	UPC bar code used for the first time when 10 Pack of Wrigley’s gum scanned
1976	Slow adoption prompts the business week article “The scanner that failed”
1978	Grocery introduces UPC
1983	Grocery completes adoption
1986	Retailers Wal-Mart and Kmart adopt UPC. Other retailers follow
1991	Wal-Mart mandates case level barcodes UCC128 by July 1992. Other retailers follow

Table 2. RFID Adoption Timeline

1940s	RFID technology invented by Harry Stockman	RFID used in WWII to distinguish friendly and foe aircraft		
1950s	D. B Harris patents radio transmission systems w/ modularly passive responder			
1960s	Commercialization of Electronic Article Surveillance			
1970s	Additional patents granted; RFID reaches consumer packaged goods	LASL releases RFID to public sector	Aimtech and Identromex formed	First implantable RFID tags used in dairy cows in Europe
1980s		Shift from performance to cost and size reduction		
1990s	Auto-ID Centre established at MIT	Los Angeles adopts pet tagging	Railroads begin use of RFID to track trains and cargo in motion	RFID-chipped Speed pass wand introduced; Gillette, P&G, UCC begin study of RFID use in theft prevention
2000s	Research and Development, Military and Government, Commercial Applications	Study of RFID use in supply chain expands from 3 to 70 corporate participants	P&G/Wal-Mart test of RFID tags for functionalit y w/in supply chain	Wal-Mart and DOD Mandates; Associated Foods Stores use RFID to track trailers

Table 3: Profile of organizations interviewed and their adoption and integration decisions
Note: * indicates those organizations that are not end users hence their responses on adoption and integration were not considered

Organization	Industry Sector	Main Supply Chain Role	RFID Adoption Role(s)	Initial Adoption	Expected Integration
A	Home Construction & equipment retailer	Retailer	End user	No	No
B	Consulting*	Solution providers	Provide expertise in RFID adoption		
C	Logistics and transportation	Logistics Support and Solution Provider	Expertise and End user	No	No
D	Label Makers And Antenna makers	Logistics	Vendors and End users	Yes	Yes
E	Reader Manufacturers*	Technology and solution providers	Vendors		
F	Beverage bottling	Suppliers	End User	Yes	No
G	Consumer products (paper based)	Suppliers	End User	Yes	Yes
H	Pallets	Suppliers	End User	Yes	Yes
I	Hardware,* software expertise	Consulting/ Solution Providers	Vendor		
J	Retail Solutions	Solution providers/manufacturing	End User	Yes	Yes

Table 4. RFID Adoption Enablers				
#	**Enabler**	**Bar Codes**	**RFID**	**Organization(s) Mentioning Enabler**
1	Standards	Consensus on Standards reached in (1973)	EPC Global Gen 2 Standards	A, B, C, D, E, G, H, I, J
2	Network effects and critical mass	Mass merchandisers adopt (1985-86)	Not yet but needed to drive tag prices down further	A, B, C, D, G, H, I, J
3	Mandates	Walmart (1991)	Walmart (2003)	A, B, C, D, E, F, G, H, I, J
4	Focus on innovation opportunity	Suppliers for grocery chains (Mid 1970's)	Suppliers for Walmart	B, C, G, H, I, J
5	Partner collaboration	Grocery Industry	Some level with Walmart but need better understanding of partner needs	B, G, I
6	Supporting Technologies	Laser and computers	Middleware and Supporting Hardware needed	B, C, D, I, J
7	Addressing consumer concerns	Consumer groups protest removal of price tags (1974)	Privacy concerns and protests on tagging	B, D,
8	Radical Innovation	Information Impact on Balance of power	Disruptive with the ability to leap frog competition and requires significant changes in business processes	A, B, C, D, E, F, G, H, I

Shifting Technological Landscape: IT Departments and Anticipated Technological Changes

Jeffrey Cummings
cummingsj@uncw.edu

Thomas Janicki
janickit@uncw.edu

Douglas Kline
klined@uncw.edu

Information Systems Operations Management Department
University of North Carolina Wilmington
Wilmington, NC 28403

Abstract

A constant in the Information Technology field is change. Technologies continue to change at a rapid pace. The need to remain current is essential for all professionals in the IS/IT field. This research presents the results of an employer survey examining the current and anticipated change in the usage of technologies. Technologies evaluated in this work include those in the areas of databases, programming languages, networking, cloud computing and operating systems platforms. Results are discussed and compared to similar surveys conducted in 2008 and 2003. While results found that Microsoft technologies are still dominant, they also suggest an increased emphasis on mobile platforms within the operating systems area and virtualization within the networking area. Within the database area, open source software (MySQL) increased dramatically compared to the prior surveys. Finally, cloud computing was included as a new technology area with mixed results.

Keywords: IT Skills, IT Technologies

1. INTRODUCTION

The demand for IT professionals continues to increase. Currently, IT-related jobs comprise four of the top ten Fortune Best Jobs of 2013 (US News, 2013). Demand is expected to continue to grow. The United States Bureau of Labor Statistics reports an expected growth of 22% in IT-related jobs for the 2010-2020 period. This outpaces most other fields through the year 2020 (US Bureau of Labor Statistics, 2011). Concurrent with the demand for general IT professionals is a need for experienced and knowledgeable professionals in a variety of technologies and skills such as databases, servers and programming languages.

The challenge facing any computer information systems professional is staying relevant. The dynamic nature of the IT field compounds the problem as employer demand for knowledgeable IT professionals continues to evolve as technologies change. This challenge is felt most in the academic community as faculty race to achieve a balance of fundamentals with relevant applied course content to meet the current and future needs of the industry (Richards et al 2011). Research has examined various skills needed by IT/IS graduates including skills within networking, project management and strategy (Janicki et al., 2004; Janicki et al., 2008). The research presented here goes beyond the curriculum to examine technology needs of current IT departments based on those employees interacting directly with the technology.

The goal of the current study is to not only identify the current changes in the usage of technologies and skills needed by employers, but also to capture anticipated changes in the coming years. Our objective is to provide an outline of technologies for current and future IT professionals to stay abreast on organizational needs. A survey was conducted to assess the needs of organization which is a follow-up study to similar surveys conducted in 2003 and 2008 (Janicki et al., 2004; Janicki et al. 2009). As such, this paper details the changes in technologies demanded by the IT industry since the prior surveys.

2. LITERATURE REVIEW

The technological environment has always been marked by frequent changes, which, in turn, requires employees to constantly adjust to meet relevant knowledge and skills needed (Lee et al. 1995). More recent technological developments ranging from virtualization to cloud computing has caused these employees to possess a blend of skills, one being a variety of technical knowledge (Byrd et al 2004). A strong background in technological knowledge is even more important in today's IT industry, as positions dealing directly with technology (e.g. network administration) are increasingly in demand with little likelihood of being "offshored" (Atkinson & Andes, 2010). This creates the challenge of developing a list of concrete technical skills needed (He & Freeman, 2010) especially as technical skills become obsolete at a much more rapid pace compared to the past (Prabhakar, et al., 2005).

Prior research has attempted to address the issue of technological needs and employer demands from a number of perspectives. One approach has been to evaluate the needs of IT management or recruiters. Sala (2011) took the approach of examining IT executives to understand their thoughts on skills in demand. The results from this research suggest programming ranks highest in demand followed project management and help desk support. Another study targeted management and above positions to develop a typology of skills and needs for an IT employee (Gallagher et al 2011). Alternatively, He and Guo (2011) focus on recruiters and their perspective on IT skills needed. However, with all this research, the focus has been on participants that may not directly interact with the technologies being surveyed. Executives and upper management as well as recruiters are not the employees directly interacting with the technologies/skills being utilized. In other words, they are not "in the trenches" of the IT department.

Another approach has been to evaluate technological need by matching the needs of the employer to the curriculum. Surendra and Denton (2009) present a comparison of skills and technologies valued by practitioners to those valued in academics. Leigler et al. (2013) examine students and their perceptions of skills needed. While some studies have focused on recent alumni (Auken et al. 2011), there are still gaps in the evaluation of experienced practitioners concerning current and anticipated skills needed.

The goal of the current study is to cover a broad range of practitioners with varying levels of experience. This paper extends the survey by Janicki et al. (2004; 2009) which longitudinally assessed the changing needs of the IT community. Specifically, we consider the current technology needs in the areas of databases, programming languages, networking, and operating systems platforms, as well as the anticipated changes in the near future. Finally, we evaluate how these needs have changed compared to the previous studies in 2003 and 2008.

3. METHODOLOGY

The survey instrument was developed over four phases represented in Figure 1. This methodology was chosen based on prior research conducting surveys in a similar manner (Janicki et al., 2004; Janicki et al. 2009).

Phase I

Phase I consisted of a roundtable discussion comprised of 25 professionals and 8 faculty from an corporate/academic advisory board. These advisory board members represent twenty unique companies ranging in size from 5 to 1000 employees. Additionally, all those on the advisory board directly interact with technology at their respective organization and were primarily members of the IS department. The goal was to develop topic areas of importance to IT professionals, specifically focusing on identifying major technology areas. Faculty representatives included members from both the Information Systems and Computer Science departments at a large, regional university.

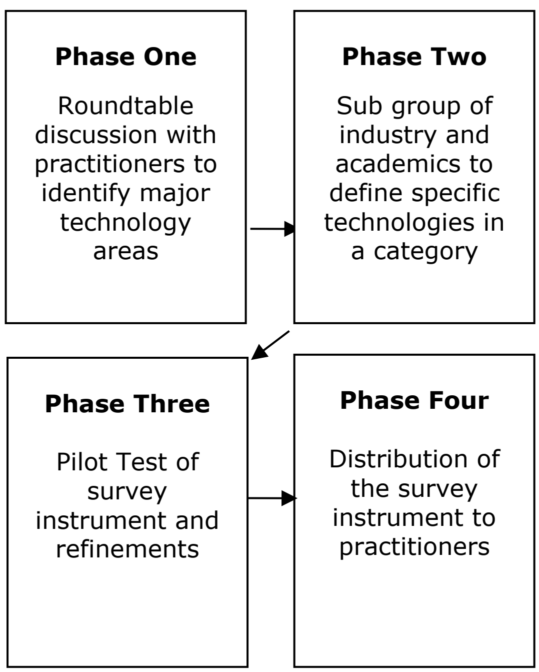


Figure 1: Survey Methodologies Stages

During the roundtable, groups were first tasked with identifying major technology areas within their field. This was then followed by an examination of the previous survey to evaluate the relevance of the technological areas previously included. The team discussions resulted in the following categories for technological areas (Note: Cloud Platforms is a new technology area added to the current survey):

- Operating Systems Platforms
- Networking/Communication (includes both Software/Hardware)
- Databases
- Development Languages
- Cloud Platforms

These overarching technology categories were then used as a starting point for phase II.

Phase II

During this phase, sub-categories of the technological areas were identified to further evaluate technological needs of employees. A sub group of the professionals from Phase I were used to define the specific items and brand names within each technology area. The sub groups went through several iterations and 'pilot testing' with other industry professionals, to ensure all possible sub-categories were captured as well as maintain consistency across areas. Appendix B provides the detail for each technology area and sub category (i.e. product or brand name). The final list of technologies/software was chosen by IT professionals based on their direct experience and thoughts as well as ongoing importance.

For example, professionals were asked to identify specific technologies and brands for the Operating System (OS) platform category. The identified technologies and brands included:

- Windows Family
- Linux/Unix Family
- MacOS
- iOS
- Android

The remainder of the survey was also developed in this phase. Since the target audience is industry professionals, questions centered on whether the technology is currently being used and what the future importance of the technology is. Due to the evolving nature of the IT field, the sub group of academics and industry professionals decided to only focus on a two year time horizon. The scale for the future importance is presented in Table 1.

Expected importance to your job in two years
Not at All
Less Important
Same
More Important
Extremely Important

Table 1: **Expected Importance Scale of Particular Technologies in two years**

There was one change to the current survey compared to the previous survey conducted in 2008. For the "Programming Languages" technology area, the "level of knowledge desired" was used to capture the current needs of the employer. This was chosen in place of "expected importance" as the sub group of employers and faculty felt it was more relevant to understand the level of knowledge needed for a specific programming language. All other technological area sub categories were asked "expected importance".

After the sub category selection was complete, the survey instrument was finalized to include questions concerning the size of the company, organization type, employee functional area and general demographics (age, gender, location, company size, industry, job title). Additionally, there were questions concerning company location and whether the participant was responsible for hiring or supervising IS/IT professionals.

Phase III

Clarity and completion time of the survey was an important aspect given that our target participant audience was professionals. A pilot test was conducted to ensure that the survey would be clear to the participants and would also have a completion time of 10 minutes or less. A preliminary survey request was emailed to twenty five industry professionals, which directed them to complete the online survey. Additional instructions were included asking all the pilot participants to record their completion time and any misleading or confusing questions. From the 25 requests, 14 completed the pilot study.

Based on pilot testing feedback, minor changes were made to the survey instrument and it was deemed ready for distribution.

Phase IV

The final phase was the distribution of the survey, via email, to over 3500 individuals either in the IS/IT field or known to potentially hire IS/IT professionals. Only those working directly on technologies were included in our survey pool so supervisors/managers were not directed to respond to the technology questions. From the remaining individuals, a total of 108 IT professionals completed the entire technology section of the survey. The survey's mailing list included the membership roster of the Association of Information Technology Professionals (AITP), alumni from the research institution, and various IT professional attendees of at least one conference held at the research institution. The goal was to poll a variety of individuals across numerous companies, geographic regions, and industries.

4. SUMMARY STATISTICS

Participants consisted of IT professionals who had direct interaction with the technologies surveyed. This included a variety of organizational roles with Software Development and Other IT leading the roles represented. A complete list of the professional roles of participants is included in Table 2.

Organizational Role	# of Respondents	%
Software Development	19	18%
Business/ Systems Analysis	11	10%
IT Strategy	11	10%
Networks/ Security	11	10%
Database Admin/ Analyst	10	9%
Management	7	7%
Project Mgmt	4	4%
Big Data / BI	3	3%
Other IT	31	29%

Table 2: Organizational Role

In the subsequent sections, the demographics are first discussed followed by the results of the importance of various technologies (by area) employed at the participant's organization.

Demographics

The participants consisted of 20% female and 80% male respondents. Education varied with a large majority of participants holding either a Bachelors of Science or Master's degree in an IT related field (30% and 19% respectively). Those with non-IT related degrees consisted of 16% with a BS and 16% with a master's degree. The overall average tenure within the industry was 15 years with employees averaging 6 years or less with their current employer.

Participants from a variety of organization types and sizes completed the survey. Over half of the participants came from organizations larger than 1000 employees and a majority identified their organization as being a Corporation. Tables 3 & 4 detail the size and type of the respondent's organization.

Number of Employees	# of Respondents	%
<11	4	4%
11-100	12	11%
101-499	24	22%
500-999	10	9%
1000-9999	31	28%
10000+	28	26%

Table 3: Size of the organizations

Organization Type	# of Respondents	%
Corporation	55	50%
Education	26	24%
Government	8	7%
Healthcare	8	7%
LLC	5	5%
Non or Not for Profit	4	4%

Table 4: Organization Type

5. RESULTS

5.1 Operating Systems Platform Expectations

OS Platform expectations were surveyed across five different platforms to understand the importance of these platforms in the next 2 years. As stated earlier, the platforms included in the survey were determined by a panel of industry and advisory board professionals. In Figures 2 through 5, the x-axis represents the number of responses.

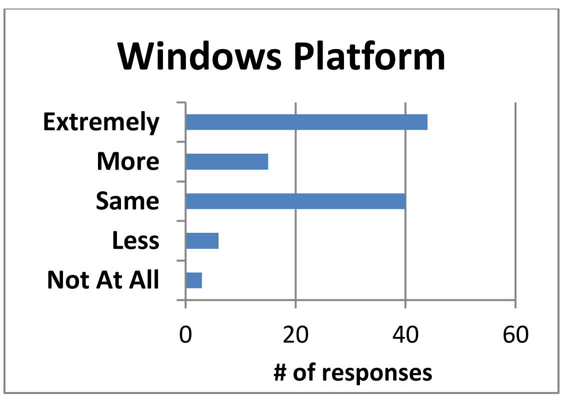


Figure 2: Expected Importance of Windows Platforms

The Windows platform was rated as the highest importance in this category. However, two new platforms introduced in this survey scored high on future importance. Both Android and iOS platforms scored the next highest ratings of importance after Windows. Figures 2 and 3 detail the top three platforms expected to have "more" or "extremely more" importance to IT professionals in the next two years. Appendix B details the responses of all platforms surveyed.

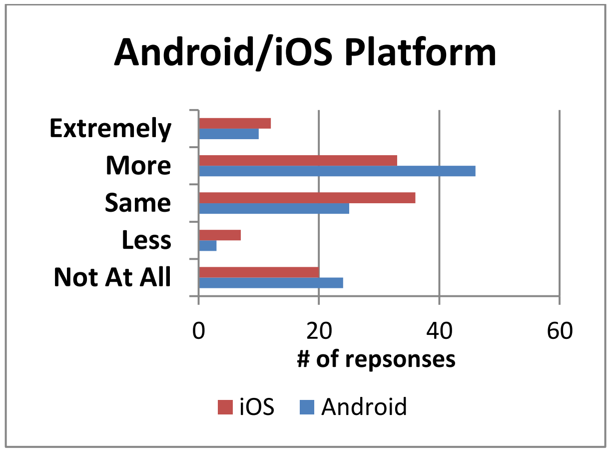


Figure 3: Expected Importance of Android & iOS Platforms

5.2 Networking/Communication

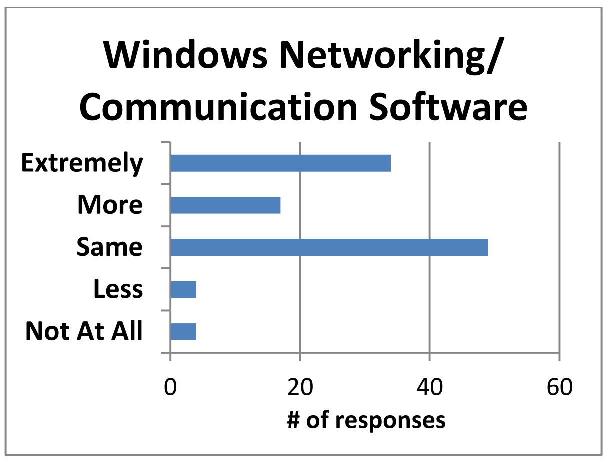


Figure 4: Expected Importance of Windows Networking/Comm. Software

The Networking and Communication category was created to include both software products (e.g. Windows Networking) and hardware products (e.g. Cisco Technologies). We included both of these to ensure we not only captured the primary software organizations use but also the popularity of as specific technology. This would help IS professionals understand the benefits of certifications in a particular technology. From this technology category, Windows networking and communications software had the highest average level of importance (detailed in Figure 4).

While Windows Networking averaged the highest ratings, Virtualization Technologies scored higher for both "More Important" and "Extremely Important" ratings than any other technology (shown in Figure 5). This shows the increasing importance placed on virtualization at organizations. Appendix B provides a complete list of the technologies surveyed included their ratings of importance.

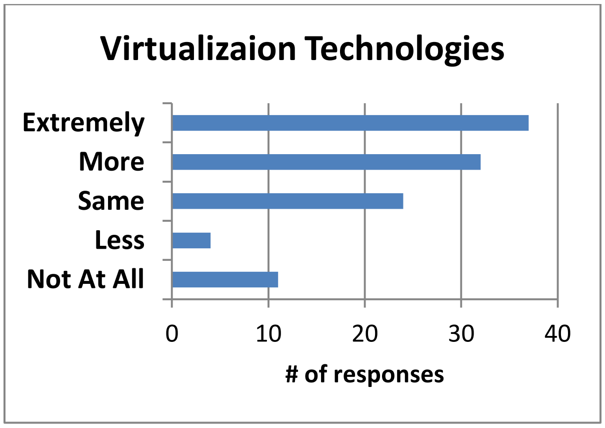


Figure 5: Expected Importance of Virtualization Technologies

5.3 Databases

MS SQL was again the leader in the database category followed by Oracle and mySQL Figure 6 provides the detailed responses for MS SQL.

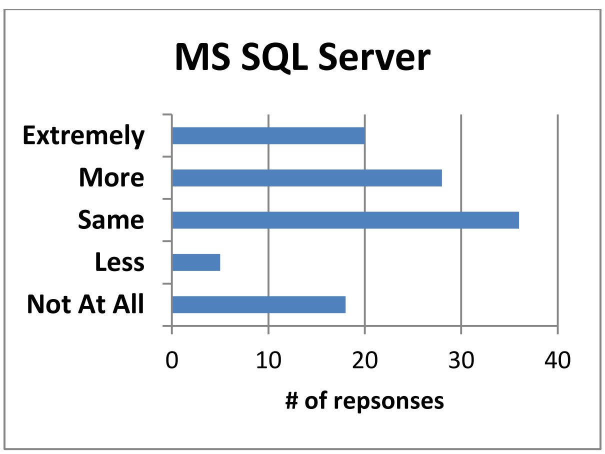


Figure 6: Expected Importance of MS SQL Server

One significant change from previous surveys was the results for mySQL. This database technology increased in importance with a large number of respondents stating this importance will remain or increase in the future.

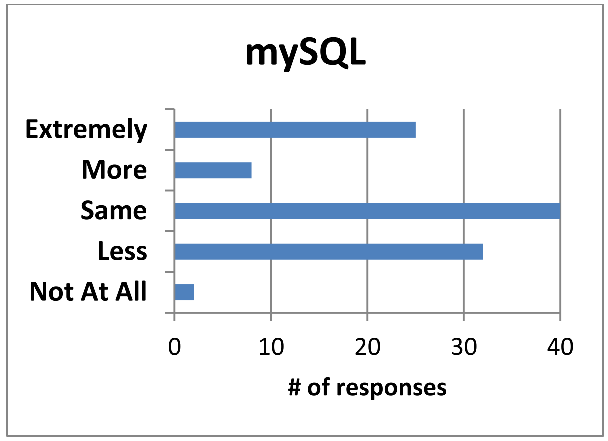


Figure 7: Expected Importance of MySQL Server

5.4 Development Languages

For Development Languages, participants were asked about (1) overall level of knowledge in development/programing and (2) desired knowledge in a specific language survey. There were a total of 9 different languages including recent languages introduced such as HTML5.

Rank	Product	Rating
1	ASP.Net	1.77
2	PHP	1.72
3	C#	1.61
4	HTML5	1.60
5	JavaScript	1.57
6	XML	1.56
7	ASP.Net MVC	1.54
8	Java	1.52
9	JQuery	1.51
10	Python	1.42
11	CSS3	1.36
12	C++	1.35
13	JSP	1.34

Table 5: Development Language Level of Knowledge Importance

The survey results found that participants indicated a need to have, at a minimum, a fundamental knowledge of software development with some suggesting working knowledge is needed as well. For the specific languages queried, ASP.Net, PHP, C# and HTML5 ranked highest of the development languages (details in Table 5). While the languages above did rank highest, the overall ratings for all languages were close, suggesting that employers are looking for individuals who have the ability to understand programming but no particular language.

5.5 Cloud Platform

A new technology area was defined for this survey that was not included in the previous surveys. Cloud Platforms were surveyed to understand the importance of using these services in the future. Surprisingly, participants responded that the expected importance of cloud platforms will remain the same or slightly less (see Table 6). These results show an average importance of 2.6 which is between the same importance in the future and less important (note: this is out of a 5 point scale with 1 being highest and 5 the lowest).

Product	Average Rating
AWS (Amazon)	2.6
FORCE	2.2
Google Web	2.9
Azure (Windows)	2.6

Table 6: Cloud Platform Rankings of Importance

A closer examination of the results shows that participants are divided concerning the importance of Cloud Services (Appendix B contains the details of responses). Across all platforms, 63% of participants rated the anticipated importance of cloud computing would remain the same or increase in importance. However, 37% of participants rated cloud services importance declining in the next two years. These results are further discussed in Section 7.

6. COMPARISON TO PRIOR SURVEYS

This research parallels prior surveys of IT workers conducted in 2008 and 2003. The number of respondents increased this year to 108 compared to 79 from the survey in 2008. These surveys were similar to the current one in context and format. However, changes were made per the suggestions of the advisory board concerning their use of technologies across those organizations.

The subsequent sections compare the changes to levels of importance across the previous surveys. All tables display the importance ranking which was calculated as follows: 5 for extremely important, 4 for more important, 3 for same, 2 for less, and 1 for not at all. A '--' indicates any products that were not surveyed in the respective year. Two of the technology areas are excluded from the comparisons. Cloud computing is not included in this section as this is a new category included in the current survey. Additionally, programming languages is excluded because of the question changes to this technology area.

6.1 Operating System Platforms

The Windows family of operating platforms has remained close to the same level of importance from previous surveys (see Table 7, higher scores indicate greater importance). However, the new products added in 2013 (i.e. iOS and Android) have the largest importance behind Windows. This suggests the increased importance companies have been placing on mobile platforms. Linux has remained consistent while Mac OS has increased significantly from 2008. Finally, the Palm and Windows CE platforms were dropped from the 2013 study.

Product	2013	2008	2003
Windows	3.8	3.9	3.9
iOS	3.1	--	--
Android	3.1	--	--
Linux/Unix	2.9	2.6	2.9
Mac OS	2.6	1.5	--
Palm	--	1.7	2.2
Windows CE	--	1.7	1.9

Table 7: Operating Platforms Rankings of Importance

6.2 Networking & Communications

All products surveyed in the networking and communications software category increased in importance from 2008 (see Table 8 for details). Virtualization and VOIP were both added to the current survey. Importance for both products was high with Virtualization tying with Windows for the highest level of importance. This may be related to organizations changing emphasis to a virtualized environment as well as security concerns that arise from these environments. Due to the decreased importance of Netware and Juniper from 2008, these products were dropped from the current survey.

Product	2013	2008	2003
Windows	3.7	3.5	3.9
Virtualization	3.7	--	--
VOIP	3.4	--	--
Wireless	3.4	3.2	3.2
Cisco	3.2	2.4	3.9
Linux/Unix	2.8	2.3	2.9
Netware	--	1.6	--
Juniper	--	1.5	--

Table 8: Networking/Communication - Rankings of Importance

6.3 Databases

For Database Products, there was an increase in importance for all products compared to the prior surveys. Microsoft SQL had the highest level of importance with MySQL and Oracle the next highest. This could be the impact of additional data analysis and the need to store more data in a variety of formats. Table 9 displays all product rankings and survey results.

Product	2013	2008	2003
MS SQL Server	3.3	3.0	3.6
MySQL	3.3	2.1	2.1
Oracle	2.8	2.7	2.9
IBM DB2	2.2	1.8	1.6
PostgresSQL	2.1	1.6	1.6
Filemaker Pro	--	1.4	1.3
CA Ingress	--	1.3	1.3

Table 9: Database Rankings of Importance

As mentioned previously, MySQL had the highest increase in importance from the previous survey. This suggests open source software may be gaining ground in organizations. This supports prior research finding the increasing popularity of open source at organizations (Zhu and Zhou, 2012).

7. CONCLUSIONS

A closer look at technologies and software at organizations has shown a dynamic landscape with some core technologies remaining important as well as newer technologies impacting the landscape. The current survey still suggests a dominance of Microsoft products across many of the technology areas surveyed. This includes the Operating Platforms (Windows), Networking/Communication, and Database (MS SQL).

However, we did see some significant shifts in a number of technology areas. While Windows still dominates Operating Platforms, mobile platforms are increasing in importance with iOS and Android leading the way. This parallels the shift seen in organizations to focus on the mobile environment.

Windows products for networking / communication were again found to have high anticipated importance. However, the importance organizations have been placing on virtualization is shown as this was ranked as important as Windows. VOIP and Wireless also saw higher anticipate importance for organizations moving forward.

Finally, the results from programming languages and cloud computing questions provided interesting insights into the anticipated importance of these technologies. There is still a high anticipated importance of knowledge in general programming/development skills moving forward. However, the results for the languages queried suggest there may not be one language that stands out. Instead, programmers need to understand the fundamentals of programming in general as well and be able to learn and adapt to the primary language used at the organization.

As previously mentioned, the results for cloud computing were mixed. While a majority of participants rated cloud service importance as remaining or increasing, we still found 37% participants rating these services as reducing in importance. This may be a result of the participant's role not being directly impacted by cloud services. Another explanation may be the employee's organization as some of the organizational types are not known to use cloud services.

8. FUTURE RESEARCH AND REMARKS

Future research includes expanding the technologies surveyed and including additional employers. While there are numerous emerging technologies (especially in areas such as cloud services), we limited the technologies in the survey to those identified by the advisory board. There are additional areas we would like to query in future surveys including social media and data analytic products. Social media adoption within the organization has potential impacts (see Cummings, 2013 for social networking adoption) so understanding the widespread use of technologies would be beneficial. Also, our data primarily came from the east coast so future research may include partnerships with organizations to include a variety of participants throughout the country.

One limitation to the survey was that the participant mailing list included a large number of supervisors/managers which were not directly asked questions on anticipated use. Future surveys will include these individuals for comparison purposes to examine what employees anticipate as important compared to their supervisors/managers. Another limitation concerned the cloud services surveyed. The survey focused on specific cloud services (e.g. vendors) without discussing cloud offerings in general. Future research is needed asking general questions concerning cloud services (software-as-a-service) to understand their importance beyond vendor specific offerings.
Lastly, while the current research focuses on technology used by IT/IS professionals, we would like to evaluate the impact of technological changes on IT curriculum. In order to meet industry expectations going forward, the academic environment needs to prepare graduates by updating the curriculum and skills of their faculty (Medlin et al., 2007). Furthermore, it touches on the question of what role should higher education play in IT skills. Should we move towards specific products and certifications, because of their high perceived value on graduates' resumes? Or do we focus on fundamental skills that are not product-specific and let organizations train graduates in the products that they use?

9. REFERENCES

Atkinson, R. D., & Andes, S. M. (2010, April 22). Looking for jobs? Look to IT [Web Memo]. The Information Technology & Innovation Foundation. Retrieved from http://www.itif.org/files/2010-wm-it-jobs.pdf on 5/27/2013

Byrd, T., Lewis, B., & Turner, D. E. (2004) The Impact of IT Personnel Skills on IS Infrastructure and Competitive IS, *Information Resources Management Journal*, 17(2), 38-62.

Cummings, J. (2013). The Impact of Intra-Organizational Social Networking Sites on Impression Formation. *Journal of Information Systems Applied Research,* 6(2) pp. 40-50.

Gallagher, K., Goles, T., Hawk, S., Simon, J., Kaiser, K. , Beath, C., & Martz, W. (2011). A Typology of Requisite Skills for Information Technology Professionals. Paper presented at the 44th Hawaii International Conference on System Sciences (HICSS).

He, J., and Freeman, L. (2010) Understanding the formation of general computer self-efficacy, Communications of the *Association for Information Systems*, 26(12), 225-244.

Janicki, T., Kline, D., Gowan, J., and Konopaske, R. (2004) Matching Employer Needs with IT Curriculum: An Exploratory Study , *Information Systems Educators Journal ,* 2004; V. 2 N. 21 ;

Janicki, T. N., Lenox, T., Logan, R., & Woratschek, C. R. (2008). Information systems/technology employer needs survey: Analysis by curriculum topic. *Information Systems Education Journal*, 6(18).

Janicki, T. N., Yaylacicegi, U., Mahar, S., & Logan, R. (2009) Anticipated changes to technologies employed by IT departments in 2009. *Journal of Information Systems Applied Research*, 2(9).

Lee, D., Trauth, E., and Farwell, D. (1995) Critical skills and knowledge requirements of IS professionals: A joint academic/industry investigation, *MIS Quarterly*, 19 (3), 313–340.

Legier, J., Woodward, B., Martin, N. (2013) Reassessing the Skills Required of Graduates of an Information Systems Program: An Updated Analysis. *Information Systems Education Journal*, 11(3), 78-89.

Medlin, B.D., S. Schneberger, and D.S. Hunsinger, 2007, "Perceived Technical Information Technology Skill Demands versus Advertised Skill Demands: An Empirical Study," Journal of Information Technology Management, 18(3-4), 14-23.

Prabhakar, B., Litecky, C., and Arnett, K. (2005) IT Skills in a Tough Job Market, *Communications of the ACM*, 48(10), 91-94.

Richards, D., Marrone, M., and Vatanasakdakul, S. (2011). What does an information systems graduate need to know? A focus on business analysts and their role in sustainability| Macquarie University ResearchOnline. *Paper presented at the ACIS 2011*.

Sala, R. (2011, September 30). 9 hot IT skills for 2012: IT hiring ticks up for the third straight year, and IT pros with programming skills will be first in line for jobs. *ComputerWorld.com.* Retrieved from http://www.computerworld.com/s/article/358381/9_Hot_Skills_for_2012?taxonomyId=14&pageNumber=1 Retrieved 5/28/2013.

Surendra, N. C., & Denton, J. W. (2009). Designing IS Curricula for Practical Relevance: Applying Baseball's" Moneyball" Theory. *Journal of Information Systems Education*, 20(1), 77-86.

United States Bureau of Labor Statistics, 2011. http://www.bls.gov/ooh/about/projections-overview.htm

US NEWS: Top Ten Jobs for 2013: http://money.usnews.com/careers/best-jobs/rankings/the-100-best-jobs). Retrieved 5/13/2013

Van Auken, S., Chrysler, E., Wells, L. G., & Simkin, M. (2011). Relating Gap Analysis Results to Information Systems Program Attitudes: The Identification of Gap Priorities and Implications. *Journal of Education for Business, 86*(6), 346-351.

Zhu, K. and Zhou, Z., 2012 "Lock-In Strategy in Software Competition: Open-Source Software vs. Proprietary Software" Information Systems Research, 23(2), 536-545.

APPENDIX A

Sample Survey Page

Please indicate what technologies you are currently using and your expected change in importance of these technologies.

PLATFORMS

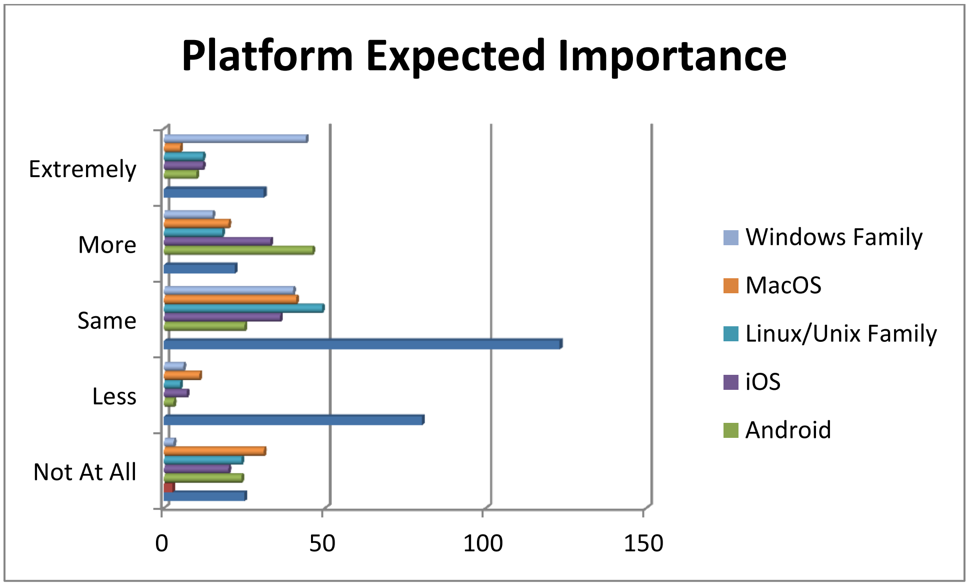
Item	Currently Used	Expected Importance to your Job in 2 Years
Windows Family	◎ Yes ◎ No	◎ Extremely ◎ More ◎ Same ◎ Less ◎ Not at All
Linus/Unix Family	◎ Yes ◎ No	◎ Extremely ◎ More ◎ Same ◎ Less ◎ Not at All
MacOS	◎ Yes ◎ No	◎ Extremely ◎ More ◎ Same ◎ Less ◎ Not at All
iOS	◎ Yes ◎ No	◎ Extremely ◎ More ◎ Same ◎ Less ◎ Not at All
Android	◎ Yes ◎ No	◎ Extremely ◎ More ◎ Same ◎ Less ◎ Not at All

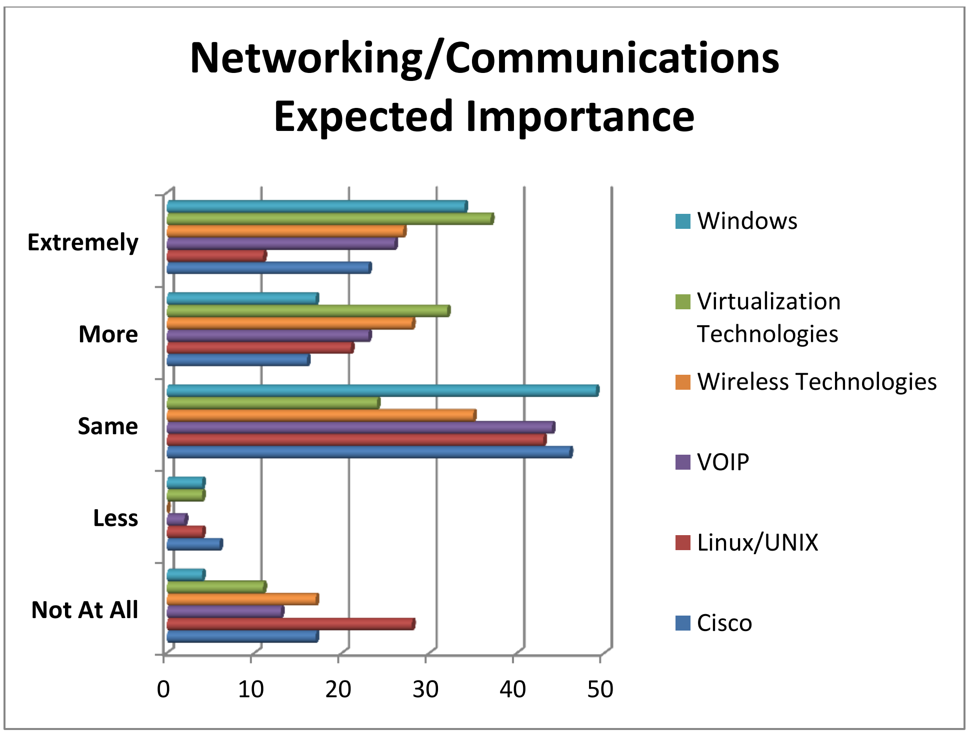
NETWORKING

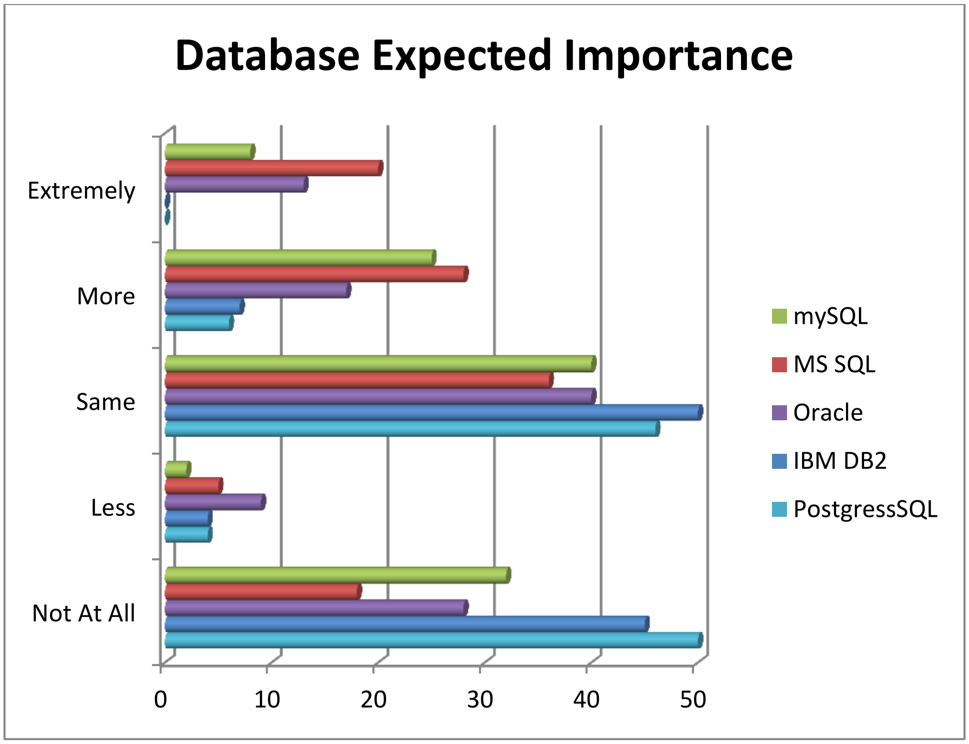
Item	Currently Used	Expected Importance to your Job in 2 Years
Windows	◎ Yes ◎ No	◎ Extremely ◎ More ◎ Same ◎ Less ◎ Not at All
Cisco	◎ Yes ◎ No	◎ Extremely ◎ More ◎ Same ◎ Less ◎ Not at All
Linux/UNIX	◎ Yes ◎ No	◎ Extremely ◎ More ◎ Same ◎ Less ◎ Not at All
Virtualization Technologies	◎ Yes ◎ No	◎ Extremely ◎ More ◎ Same ◎ Less ◎ Not at All
VOIP	◎ Yes ◎ No	◎ Extremely ◎ More ◎ Same ◎ Less ◎ Not at All

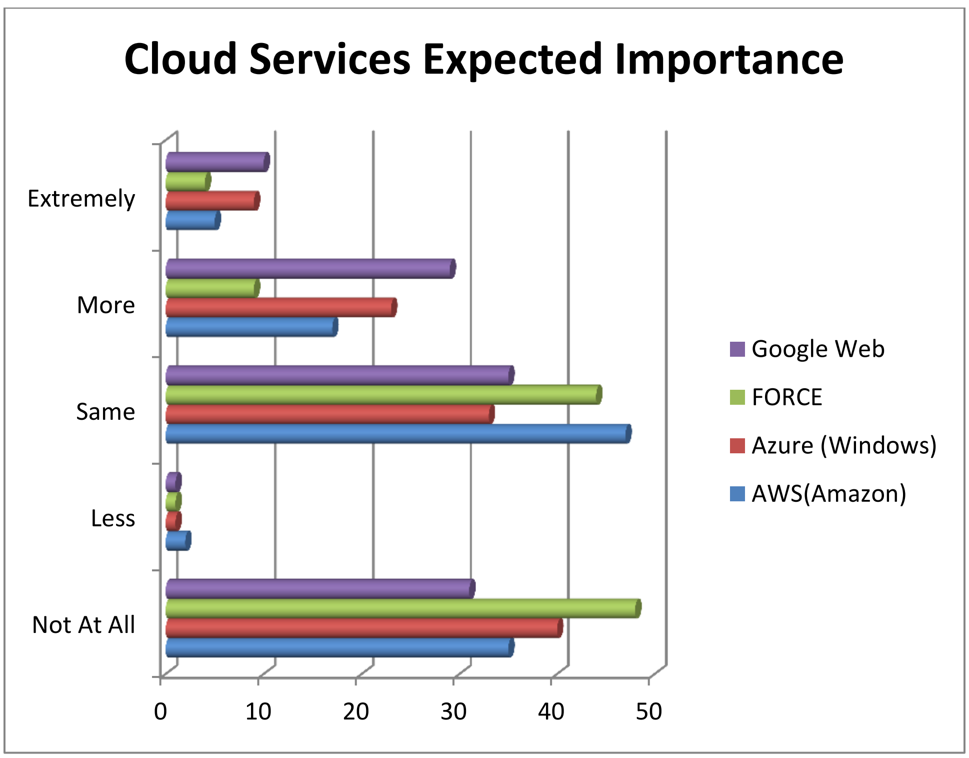
APPENDIX B

Technology Area Survey Results for Expected Importance


Platform Expected Importance
Extremely
More
Same
Less
Not At All
0
50
100
150
Windows Family
MacOS
Linux/Unix Family
iOS
Android


Networking/Communications Expected Importance
Extremely
More
Same
Less
Not At All
0
10
20
30
40
50
Windows
Virtualization Technologies
Wireless Technologies
VOIP
Linux/UNIX
Cisco


Database Expected Importance
Extremely
More
Same
Less
Not At All
0
10
20
30
40
50
mySQL
MS SQL
Oracle
IBM DB2
PostgressSQL


Cloud Services Expected Importance
Extremely
More
Same
Less
Not At All
0
10
20
30
40
50
Google Web
FORCE
Azure (Windows)
AWS(Amazon)

Risk Assessment & Management in Merchant Capture Systems: A Threat Analysis Perspective

Kevin Streff
kevin.streff@dsu.edu

Sarin Shrestha
sshrestha17600@pluto.dsu.edu

Cody Delzer
cj.delzer@pluto.dsu.edu

College of Business & Information Systems
Dakota State University
Madison, SD 57042, USA

Abstract

Merchant Capture Systems (MCS) provide the ability to deposit checks remotely without visiting a brick-and-mortar bank. The adoption of this technology is increasing rapidly; however, security threats exist with merchant capture systems. This paper examined two prominent merchant capture architectures to determine and prioritize common security threats and mitigating controls. Threats were identified for three components of a typical merchant capture system: bank, merchant and technology service provider. The paper communicates common MCS threats and controls as gathered by a questionnaire, evaluated by security experts and verified by IT auditors and bank examiners. The study determined the likelihood and impact of each threat, calculated an asset threat score and an inherent risk score for a merchant capture system, and concluded data loss as the top security risks when checks are deposited remotely through a merchant capture system.

Keywords: Risk Assessment, Risk Management, Remote Deposit Capture (RDC), Merchant Capture System (MCS), Banking industry, and Security Threats.

1. INTRODUCTION

Remote Deposit Capture (RDC) emerged to automate the check deposit process, allowing business customers to remotely scan checks and transmit the scanned image to their financial institution without physically delivering the check to the depositary bank (Levitin, 2009). Merchant capture systems (MCS) are forecasted as a technology banks will most likely implement over the next several years, with an estimated 5 million capture points by 2014 (Meara, 2008) and 7.3 million users by 2015 (MarketsandMarkets, 2010).

The term "merchant" refers to business clients, such as the retailers, car dealers and other types of commercial clients, who desire remotely deposit checks without visiting the bank. Merchant capture systems appeared in the digital economy as a new method of providing commercial clients flexibility to deposit checks from remote locations. For example, a Wal-Mart store may accept thousands of checks daily which need to get deposited as soon as possible. Wal-Mart would enjoy the benefits of no longer having to travel to the bank to deposit checks, reducing costs and floats. However, these efficiencies potentially expose banks to information security risks (McLaughlin, 2008).

The use of merchant capture became popular after the Check 21 Act was passed (Check21, 2003; Giudice & Johns, 2009) that mandated that digital checks become the legal equivalent to physical checks: that banks must accept a digital check just as they accept a paper check. MCS creates a digital image of the original check and sends the image to a financial institution for deposit via the Internet as an encrypted file (Fisher, 2009; Levitin, 2009).

Every financial institution must identify and mitigate the security and privacy issues before implementing MCS (FDIC, 2009). A comprehensive risk assessment is required to identify threats to MCS so management understands and addresses information risks. This paper examines MCS threats for three threat actors involved in a typical MCS: banks, merchants and technology service providers (TSP). Compensating controls are discussed to mitigate each identified threat.

2. ARCHITECTURES OF THE MERCHANT CAPTURE SYSTEMS

A MCS requires a PC, application software, an Internet connection, a check scanner and optionally a technology service provider. Scanners are typically sent to merchants via the postal service. Upon receipt, the merchant installs the scanner with telephone support. Yet other providers include the user manuals together with the scanners, so clients can install them (Valentine, 2008). Merchants will reserve a PC and load it with a MCS application that will process the deposit information. Based upon different types of MCS, these applications can be either software directly installed on their computers or website accessed designed to process depositing information.

Large banks typically develop their MCS solution while community banks generally license off-the-shelf solutions (Houseman & Nevle, 2009). The two common architectures of the MCS are: Merchant Capture with TSP and Merchant Capture without TSP.

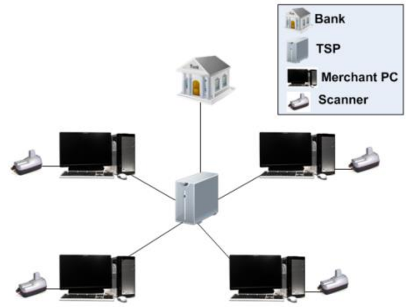


Figure 1: Merchant Capture System with TSP

Figure 1 demonstrates a typical MCS with a TSP. First, a merchant operator/employee scans the paper checks. As the checks are scanned, an image of each of check is generated and displayed on the PC, and the operator manually inputs the amount of money for each check. Alternatively, the PC may be provided with optical character recognition (OCR) software that is adapted to obtain the dollar amount of each check directly from the scanned image. In this case, the operator views the check images and verifies the amounts that are recognized by the OCR software. Once the images are created and the value of each check is obtained, the operator is asked to input the information including the account number of the client company to receive the deposit and some other information used to verify the sender's identity, such as the name of the operator, the company's address and the telephone number. The final step is a transfer of the data to the TSP's server, which analyzes the image quality of the scanned checks and forwards the deposit information through the Internet to the depository bank (Forth, Pierce, & Carey-Steckbaucer, 2007).

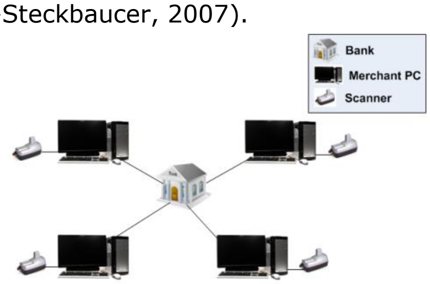


Figure 2: Merchant Capture System without TSP

Figure 2 illustrates a MCS without the service provider involved. Most of the detailed operational procedures remain the same as the former architecture; however, the deposit information from merchants is sent directly to the financial institution via the Internet.

3. BACKGROUND

Although remote deposit capture is a relatively new technology, remote banking can be traced decades back. Designed for serving customers without having to visit a bank personally, remote banking was initiated in the late 1960s with the introduction of an Automated Clearing House (ACH) system that made Electronic Funds Transfer (EFT) available. The first Automatic Teller Machine (ATM) was installed at a Chemical Bank in New York in 1969 (Bielski, 2008). ATMs introduced new conveniences, leading to a boom in ATM usage in banks, shopping malls, grocery stores, airports and other places of convenience. 403,000 ATMS were in use throughout the United States with average 239 new ATMs installed per day worldwide (Gammon, 2009).

Remote deposit platforms were extended to phone-based models in 1989 when First Direct, the first telephone based bank, was launched in the UK by HSBC bank (FinancialNews, 2008). This new concept of banking let the customers open accounts, make transactions, buy stocks and pay bills through their telephone system. Phone banking provided several advantages, including eliminating the cost of building new branches (Lennon, 1996).

Internet banking first appeared in the mid 1990s. Security First National Bank was the first bank offering Internet banking in 1995. Through the World Wide Web, customers browse their account information, carry out transactions and track payments on their PC. Internet banking was quickly offered by thousands of banks around the world by 1997. Today, most banks around the globe have a website to serve customers remotely through the Internet (Chou & Chou, 2000), yet few system include a remote deposit capability.

Mobile banking was conceptualized in the late 1990s to conduct banking commerce using mobile phones. However, technology concerns including screens which lacked the capability to show precise information, low-speed mobile phone networks, delayed adoption until 2000. With the advancement of high-speed networks for mobile phone, such as EDGE, GPRS, and 3G networks, mobile banking became practical (Riivari, 2005) to remotely deposit from a mobile platform.

To encourage innovation and efficiency in the payment system, Check Clearing for 21st Century Act (Check 21) became effective in 2004 (Check21, 2003; Jesus, 2006) to allow financial institutions to create, transmit, deposit and utilize an electronic image of the original check. Instead of transporting the original check to the bank, checks are cleared based upon a digital image. RDC models were introduced to scan and send electronic digital documents of deposit information from various remote locations (FDIC, 2009). When the digital documents are ready, they are sent to the financial institutions to complete the deposit process using specialized software (Fisher, 2009; Levitin, 2009).

According to the American Bankers Association (ABA) Banking Journal, merchant capture systems are very popular. Nearly 65% of American banks offer RDC; 58.5% of banks report that offering RDC does attract new business clients and 71.2% of banks state improved business client retention because of RDC (ABA, 2007, 2008). RDC customers are growing by 45 new customers per week, and more than 50% of the total commercial deposits are gathered through RDC (AmericanBanker, 2008). The number of scanners deployed for remote capture in the United States exceeded 700,000 scanners in 2011 (Meara, 2011). Celent's 2008 State of RDC report states two-thirds of U.S. banks have adopted the technology by the end of 2008 (Meara, 2008). Aite Group estimates that 350,000 accounts are enabled with RDC capability (Aite, 2010). Large banks leverage RDC to expand geographically, whereas small banks use RDC to substitute building physical branches (Ginovsky, 2008). Many credit unions utilize RDC to enhance their operational efficiency and reduce the courier costs of transporting paper check (Johnson, 2009). Bob Meara, a senior analyst in the banking group at Celent LLC, stated that, "In the history of U.S. financial services, there has never been a technology adopted faster than RDC" (Celent, 2008), with 7,100 financial institutions offering at least one commercial RDC solution by the end of 2011 (Chilingerian, 2011).

The Financial Crimes Enforcement Network identified 1,017 suspicious activity reports for violations pertaining to the remote capture (Bishop, 2011). American Banking Association's 2011 Deposit Account Fraud Survey estimated $893 million check related losses in 2010 (ABA, 2011). According to 2012 Faces of Fraud Survey, check fraud is listed in the top security threat (BankInfoSecurity, 2012). These circumstances support the need to risk assess and manage MCS at every bank.

4. RISK ASSESSMENT AND MANAGEMENT

Security risk management is a continuing process of identifying and prioritizing risks to minimize, monitor and control the probability and impact of unfortunate events (Spears & Barki, 2010). Various risk assessment models have been proposed. National Institute of Standards and Technology (NIST, 2010) proposed a framework in NIST SP 800-37 to improve the information security posture, and reinforce risk assessment processes to encourage cooperation among federal organizations. Saleh, Refai, and Mashhour (2011) proposed a risk assessment framework that discovers system's threats and vulnerabilities.

Similarly, numerous information security standards and guidelines have been proposed and developed to protect information assets. Gramm Leach Bliley Act is a federal law for financial institutions to develop, implement, and maintain administrative, technical, and physical safeguards to protect the security, integrity and confidentiality of information (FDIC, 2001; GLBA, 1999). Generally Accepted System Security Principles is a mutual effort to develop and maintain a set of rules, practices, and procedures to achieve information confidentiality, integrity, and availability between international countries, unifying and intensifying upon existing authoritative sources (Grimaila & Kim, 2001; Poore, 1999). The Federal Information Processing Standards is issued by the National Institute of Standards and Technology (NIST) to provide mandatory guidelines such as for security and interoperability for government agencies (FIPSPUBS, 1996).

Innovative models have been explored and deployed; however, due to the number of calculations that are performed when conducting a risk assessment, it is common for banks to employ a one-to-one-to-one or one-to-one-to-few method, which leads to the assumption that one asset has one threat and that one threat has either one or a few controls to mitigate the risk imposed by the threat. However, a systematic and accurate risk assessment method uses a one-to-many-to-many approach. This method assumes that each asset has many threats, and each of those threats has many controls to mitigate the risk.

Podhradsky, Streff, Pauli, and Engebretson (2011) conceptualized an automated risk assessment model which allows a bank to complete comprehensive and thorough one-to-many-to-many risk assessments. This method would define generic assets, each with a unique protection profile. The method would allow banks to develop protection profiles based upon the confidentiality (C), integrity (I), availability (A), and volume (V) of data each asset processes, stores, and transmits; identify threats based upon their impact and likelihood; apply controls, and generate risk reports. When the remote deposit is considered, the bank also has to focus on technology controls, such as network security settings, controls over the transaction, encryption, and physical security controls (Joseph, 2011).

5. RESEARCH METHODOLOGY

In this study, merchant capture systems were studied to understand protection profiles, threats and mitigating controls. Regarding protection profiles, the study leveraged the Podhradsky et al. (2011) approach and assigned the merchant capture system asset a protection profile (APP) based upon a high, medium, or low categorization. These qualitative ratings are turned into quantitative ratings of 3, 2, and 1 respectively. Hence, the asset MCS has been assigned an Asset Protection Profile rating of 9 of high confidentiality, high integrity, medium availability and low volume.

Confidentiality (C)	High	3
Integrity (I)	High	3
Availability (A)	Medium	2
Volume (V)	Low	1
Asset Protection Profile (APP)		**9**

Table 1: MCS Asset Protection Profile

High confidentiality as information is sensitive; its disclosure would violate federal banking regulations and/or result in significant harm to the institution. High integrity as accuracy of the information is critical; its modification or incorrectness would cause significant issues. Information availability is of moderate concern; recovery must be made within few days. Volume is rated low as only a small amount of information is regularly stored, processed, or transmitted using merchant capture systems. The CIA-V rankings are described further in Table 6 of the Appendix.

The research team developed a model that allows for reliable risk assessment improvement through identification of threats, and a process for verification and adjustments to the formerly identified threats. The major components in this process include:
• Researchers (One Professor and two Graduate Assistants).
• Security Consultants (content team composed of four InfoSec consultants who generate MCS risk assessment based upon results of the Survey Questionnaire).
• Risk Assessment Database (Captured research results for MCS threats and controls).
• Third-party IT auditors (Audit the MCS risk assessment and provide feedback).
• Bank Examiners (Examine the MCS risk assessment and MCS audit to provide feedback).

The complete process consists of seven steps is illustrated in Figure 3.

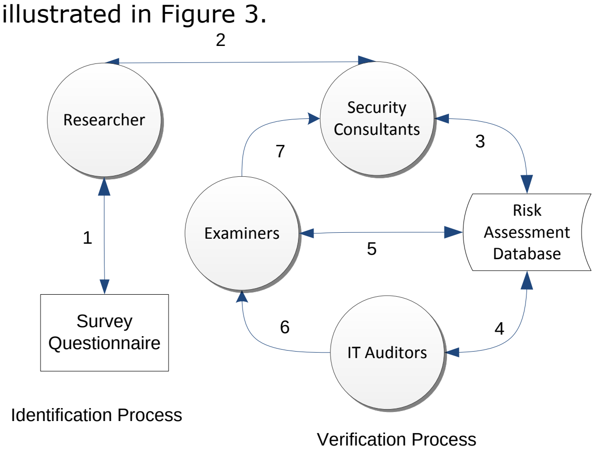


Figure 3: Research Methodology

The methodology can be sub-divided into two separate sub-processes: Identification and Verification. Identification is the process where researchers ask MCS users and experts what they believe the typical threats and controls are for merchant capture systems. Verification is the process where researchers refine the data gathered during Identification to improve the quality of the data.

Identification: The sub-process on the left of Figure 3 is intended to be a process of threat identification and initiates with Researchers collecting data from survey questionnaire that identify potential threats to MCS systems. The questionnaire was distributed among 125 individuals who were exposed in this MCS processes. Among them, 98 responses were collected. These individuals were bankers, merchant operators, and information security officers who completed the survey individually and the results were tabulated. The threats were listed with the columns of impact and likelihood with the basic rating of high, medium and low in the table for all three actors. The survey questionnaire is shown in Appendix.

Verification: After Identification is complete, a group of security consultants conducts a MCS IT Risk Assessment and provides the results in the Risk Assessment Database. The security consultants used the data collected via the survey questionnaire to develop the MCS risk assessment in ten banks. Next, a group of IT Auditors reviewed the data inserted by security consultants and provide their own recommendations to threats identified and Asset Protection Profiles and inserted the refined data into the Risk Assessment Database. Finally, a group of bank examiners, who identify improvement opportunities for risk assessment, examines the data and provide their own perspectives to identified threats and Asset Protection Profiles and inserted the data into the Risk Assessment Database. The process completes its first iteration with step 7, when the group of security consultants review findings and recommendations from independent third party IT auditors, and finally return with that information to step 3 where the process of improvement continues to cycle, ensuring the Risk Assessment database remains accurate and relevant.

Each threat is assigned a threat score called Asset Threat Score (ATS) based on its Impact (I) and Likelihood (L). Impact is the estimated degree of loss or damage to the asset or institution and assigned a value between 1 and 3 with 1 representing little or no impact and 3 representing devastating impact. Likelihood is the estimated degree of possibility the threat may have an impact on the organization in a given time and assigned a value between 1 and 3 with 3 representing that it is highly likely that a threat will occur and 1 vice-versa. The impact and likelihood rating for each threat is multiplied to produce an Asset Threat Score. Finally, the Total Asset Threat Score (TATS) is equal to the sum of all asset threat scores and the Asset Inherent Risk Score (AIRS) is calculated by taking the Asset Total Threat Score times the

Asset Protection Profile. These concepts are defined in Table 5 in Appendix.

6. RESULTS AND DISCUSSIONS

Using the methodology described above, the top MCS critical threats are described for banks, merchants, and service providers.

MCS Critical Threats for Banks

MCS threats for the Banks include:

	Bank		
Threats	I	P	I*P
Data Loss	H	M	6
Outsourced	H	M	6
Unauthorized Physical Access	H	M	6
Unauthorized System Access	H	M	6
Degraded/ Unavailable	M	M	4
Eavesdropping/ Sniffing	M	M	4
Hardware Failure	M	M	4
Intentional Misuse	M	M	4
Malicious Software	M	M	4
User Error	M	M	4
Unauthorized Remote Access	H	L	3
Environmental Incident	M	L	2
Man-made/ Natural Disaster	M	L	2
Software Acquisition	M	L	2
Social Engineering	M	L	2
Unauthorized Viewing	L	L	1
Total Asset Threat Score			**60**
Asset Inherent Risk Score			**540**
Legend:			
"I"—Impact; "P"—Likelihood; "; "H"—High; "M"—Medium; "L"—Low			
Value:			
High = 3; Medium = 2; Low = 1			

Table 2: Bank MCS Threats

Data Loss
Data loss is the insider risk of stealing or providing unauthorized access to sensitive information. Data Loss Prevention (DLP) systems have become common place in large banks while small and medium-sized banks generally find them unaffordable. Other solutions to prevent data loss include employee operational and security training, and well-composed security policies and procedures.

Unauthorized Access
Unauthorized physical and system access are significant risks to the bank in MCS. By the physical method, the attacker can steal or sabotage the bank's physical assets in the MCS while by the system method, the attacker can steal or delete the information in the MCS. If the attacker gains access to the administrator account, he/she can change the security settings in order to install malicious software that creates backdoor to the system. Although the impact of unauthorized access is obviously high, likelihood of occurrence remains medium for each of the methods since the bank usually has sound physical security considerations.

Outsourced
According to statistics from the ABA Journal, half of the banks among the U.S. choose outsourced solutions (ABA, 2007). It is true that having a third party involved in the process of MCS can remove many burdens from the bank, such as MCS infrastructure development, signing multiple sourcing contracts, updating and maintaining the software and hardware, and customer trainings (Houseman & Nevle, 2009). However, a new threat called outsourced is generated under the situation. This threat can be mitigated by performing due diligence in selecting service providers; it still has high impact.

MCS Critical Threats for Merchants

The most critical threats to merchants are found in Table 3. These threats are related in that they are associated with people. Some of them including data loss, unauthorized physical, and system access have already been described from the perspective of bank. Threats like social engineering and intentional misuse demonstrate that people the most risky part for the merchant. Unlike employees from banks paid by monthly salaries, many small business employees are seasonal or hourly. The requirement of the educational background at a merchant site is likely less than the educational requirement at a bank. Most of the merchant companies would not do employee background checking since they are hiring part-timers. The low education requirement and the lack of employee background checking leave the potential that people with a criminal background or poor credit histories might get the job which increases the probability of high intentional misuse.

Threats	Merchant I	P	I*P
Unauthorized Physical Access	H	H	9
Data Loss	M	H	6
Intentional Misuse	M	H	6
Social Engineering	M	H	6
Unauthorized System Access	M	H	6
Eavesdropping/ Sniffing	M	M	4
User Error	M	M	4
Malicious Software	H	L	3
Degraded/Unavailable	L	M	2
Hardware Failure	L	M	2
Environmental Incident	L	L	1
Man-made/ Natural Disaster	L	L	1
Unauthorized Viewing	L	L	1
Total Asset Threat Score			**51**
Asset Inherent Risk Score			**459**
Legend:			
"I"—Impact; "P"—Likelihood; "; "H"—High; "M"—Medium; "L"—Low			
Value:			
High = 3; Medium = 2; Low = 1			

Table 3: Merchants MCS Threats

Most of the merchant companies do not have security policies, procedures and do not offer security training. The consequence is increasing the probability of data loss and social engineering. In addition, not all merchants employ layered security like a bank, which means assets like checks, computers, and scanners are not hard to access by criminals.

MCS Critical Threats for TSPs

In MCS, a TSP typically deals with providing software and hardware, customer training, system maintenance, and data manipulating, which includes gathering data from the merchant then transferring it to the bank. The data from merchants being transferred to the TSP is considered sensitive, due to the data containing the image of the scanned checks to be deposited. Given this fact, data loss is possible for the employees of the TSP, who have access to the data, either by unintentionally exposing the data to the public or deliberately selling it. Although, a data breach from insiders can be mitigated by employee training and background checking, there are still many external attackers targeting TSPs.

MCS threats for the TSP include:

Threats	Third Party I	P	I*P
Data Loss	H	H	9
Malicious Software	H	H	9
Social Engineering	H	H	9
Unauthorized System Access	H	M	6
Degraded/ Unavailable	M	M	4
Eavesdropping/ Sniffing	M	M	4
Hardware Failure	M	M	4
Intentional Misuse	M	M	4
Unauthorized Physical Access	M	M	4
Environmental Incident	M	L	2
User Error	M	L	2
Man-made/ Natural Disaster	M	L	2
Software Acquisition	M	L	2
Unauthorized Remote Access	M	L	2
Unauthorized Viewing	L	L	1
Total Asset Threat Score			**64**
Asset Inherent Risk Score			**576**
Legend:			
"I"—Impact; "P"—Likelihood; "; "H"—High; "M"—Medium; "L"—Low			
Value:			
High = 3; Medium = 2; Low = 1			

Table 4: TSP MCS Threats

Malicious software like Trojan horses and backdoors are good weapons for attackers to gain access to data from the computer systems of the TSP. Combined with a little social engineering like email spam, malicious software can be installed on the system by innocent employees who fall into the trap. Another option an attacker may use is to hire someone from the TSP to install the software directly onto the system. Therefore, data loss, malicious software, and social engineering are the three most critical threats for the TSPs.

Top MCS Critical Threats for Banks, Merchants and TSP

The summary of critical threats to the Bank, Merchant and Third Party is in Table 7 in Appendix. The top five MCS critical threats are identified for banks, merchants, and service providers are as:

Data Loss
As one of the most critical threats, data loss can be described as someone intentionally or unintentionally releasing information to unauthorized recipients. The threat can be conducted either by not-well trained or by disgruntled employees through sending sensitive information of the company to unauthorized individuals or posting the information directly on the Internet. In March 2010, a former employee with TD Bank releases the customer information to accomplices who withdrew more than $200,000 from 13 bank customer accounts (Patel, 2010).

Malicious Software
It is a program that performs unauthorized processes that will lead to a malicious impact on the information systems confidentiality, integrity and availability. Symantec Corporation discovered more than 240 million distinct new malicious programs in 2009, a 100% increase over 2008 (Symantec, 2009).

Social Engineering
Rather than using complex computer techniques to gather information from the target system, social engineering is an attack based on deceiving users or administrators by an unauthorized person masquerading as a rightful user. This attack is usually performed in an attempt to gain illicit access to systems or confidential information. A mobile banking application on Android platform in December 2009 caused more than 50 fraudulent banking applications to appear (Patel, 2010). These applications attempted users to enter their bank account numbers, password and other personal information.

Unauthorized Physical Access
It is defined as someone intentionally infiltrating a secure area. It can be exploited by external attackers or internal disgruntled employees. The consequences of the threat range from theft, sabotage, even to unauthorized system access.

Unauthorized System Access
It is gaining unauthorized access to a system by physically interacting with it. The example of exploiting this threat is where an unauthorized individual or an attacker logs into the system with stolen credentials or bypassing security through hardware using CD drives and USB ports.

Top MCS Controls for Threats Identified for Banks, Merchants and TSPs
The financial institution should assess potential risks and regulatory constraints under Bank Secrecy Act when implementing MCS (FDIC, 2009). There is a healthy relationship between financial institutions and a payment processor. A payment processor is a customer who deposits checks and process payments for third party merchant clients (FDIC, 2012). Usually, payment processors effects legal payment transaction for merchant clients, the risk profile can deviate depending on the customer type. For instance, payment processors that deal with online clients may have a high risk-profile as they have the tendency to display a higher prevalence of illegal activities or fraud when compare to other businesses. Financial organization should comprehend, authenticate, and examine the activities and the entities associated to the account relationship and also outline the comprehensible lines of responsibility for governing risks related with the payment processor relationships (FDIC, 2012). The control for mitigation includes inspection and monitoring of accounts for suspicious activity, enhanced due diligence and consumer complaints. Implementing proper countermeasure may facilitate to discover payment processors that process items for fraudulent or unscrupulous merchants. To limit the potential risk associated, the financial institutions should implement risk mitigation policies including appropriate controls for the risk and procedures designed to reduce the probability of unauthorized transactions used by unscrupulous merchants.

Top controls to mitigate those threats include:

Controls for Threat- Data Loss
i. Security Information and Event Management: An application that collects, stores and analyzes security log data from multiple systems for data retention and detection of unauthorized activity.
ii. Unique User Accounts: The process of assigning unique usernames that allow them to distinguish from each other and prevent them from being guessed easily.

iii. User Activity Logs: Logs used to track details about transactions and events performed by a user.
iv. Administrator Activity Logs: Logs used to track details about transactions and events performed by an administrator.
v. Data Loss Prevention: A tool that actively monitors, blocks, and reports on data leaving the Bank to ensure sensitive information is not transmitted to unauthorized parties. Data is allowed or blocked based on the analysis of its content, instead of its source or other criteria.

Controls for Threat- Unauthorized Physical Access
i. Remote Capture User Security Controls Audit: Performing a review of the remote user's security controls to ensure the system is sufficiently protected.
Examples of a Remote Capture User Security Controls Audit are On-site, Self-Assessments, Independent Reviews, etc.
ii. Restricted Access Area: A secured area accessible only by authorized personnel or by those granted temporary access.
iii. Surveillance Cameras: Cameras that provide archived surveillance for an area.
iv. Monitored Location: Locating an asset where it will be visible by an employee who is responsible for its physical security.
v. Motion Detection: A system that triggers an alarm or other event when it detects motion.

Controls for Threat- Unauthorized System Access
i. Security Information and Event Management
ii. Remote Capture User Security Controls Audit
iii. Unique User Accounts
iv. User Activity Logs
v. Firewall -Ingress Filtering: A dedicated appliance or software running on individual computers that inspects network traffic passing into the network and denies or permits passage based on a set of rules.

Controls for threat- Social Engineering
i. Data Loss Prevention
ii. Incident Response Program: Actions to be taken when the institution suspects or detects that unauthorized individuals have gained access to customer information systems, including appropriate reports to regulatory and law enforcement agencies. The goal of an incident response program is to minimize damage to the institution and to its customers through intrusion containment and the restoration of systems.
iii. Inactive Lockout: Locking a user's session after a specified period of inactivity.
iv. Website Filtering: Prevents computer users from viewing inappropriate or unauthorized websites.
v. Social Engineering Security Awareness: Educating employees on identifying and preventing social engineering attempts.

Controls for threat- Malicious Software
i. Security Information and Event Management
ii. Firewall -Ingress Filtering
iii. Intrusion Detection / Prevention: A security management system to identify and prevent possible security breaches, which include both intrusions (attacks from outside) and misuse (attacks from within).
iv. Back-up Critical Data: Completing regularly scheduled backups of critical information.
v. Formal Patching Process: A defined process for identifying missing updates and patches and deploying them on a scheduled basis or immediately if needed.

Table 8 in Appendix illustrates the top five controls for each identified threat above.

7. CONCLUSION

The adoption of Merchant Capture System poses challenges and security threats to financial institutions. Hence, prior implementing this system, a risk assessment should be performed to identify risk and security threats. The paper identifies the most common threats and controls for MCS to support the risk assessment process at a bank. Management should also ensure the appropriate policies and controls are in place to mitigate those threats including physical and logical access controls over Merchant Capture System.

8. REFERENCES

ABA. (2007). Remote Deposit Capture: Bucks into bits begins to bite. *ABA Banking Journal, 99*(3), S4-S6.

ABA. (2008). Remote Deposit Capture: Making The Promise Pay Off. *ABA Banking Journal, 100*(3), S11-S18.

ABA. (2011). 2011 Deposit Account Fraud Survey: American Banking Association.

Aite. (2010). Remote Deposit Capture Risk Management: Raising the Bar (Vol. 166): Aite Group.

AmericanBanker. (2008). Remote Deposit Capture Partnership for Success. *American Banker Magazine, 118,* 11.

BankInfoSecurity. (2012). 2012 Faces of Fraud survey. BankInfoSecurity: Information Security Media Group.

Bielski, L. (2008). Eight Tech Innovations that took banking into the 21st Century. *ABA Banking Journal, 100*(11), 84-89.

Bishop, B. (2011). The SAR Activity Review - Trends Tips & Reviews (Vol. 20, pp. 1-94). http://www.fincen.gov/news_room/rp/files/sar_tti_20.pdf: Financial Crimes Enforcement Network.

Celent. (2008). Remote deposit capture carries risk as well as convenience. In M. Savage (Ed.).

Check21. (2003). Public Law 108-100: An Act (pp. 117). 108th Congress.

Chilingerian, N. (2011). RDC Reaches Adoption Plateau *Issue of Credit Union Times*: Celent.

Chou, D. C., & Chou, A. Y. (2000). A Guide to the Internet Revolution in Banking. *Information Systems Management, 17*(2), 51.

FDIC. (2001). Examination Procedures to Evaluate Compliance with the Guidelines to Safeguard Customer Information.

FDIC. (2009). Risk Management of Remote Deposit Capture from http://www.ffiec.gov/pdf/pr011409_rdc_guidance.pdf

FDIC. (2012). Payment Processor Relationships Revised Guidance: Federal Deposit Insurance Corporation.

FinancialNews. (2008). History of the First Direct Bank

FIPSPUBS. (1996). Federal Information Processing Standards Publications Retrieved 04/07, 2013, from http://www.itl.nist.gov/fipspubs/geninfo.htm

Fisher, D. M. (2009). Exam Council spells out remote capture risks. *ABA Banking Journal, 101*(3), 34-35.

Forth, T. J., Pierce, J. D., & Carey-Steckbaucer, H. (2007). USA Patent No. 20070086642. P. B. Inc.

Gammon, K. (2009). ATMs by the Numbers. *Wired Magazine*.

Ginovsky, J. (2008). Reviving up for Remote Deposit Capture. *Community Banker, 17*(9), 28-30.

Giudice, E. L. D., & Johns, N. Z. (2009). Risk Management for Remote Deposit Capture. *Bank Accounting & Finance, 22*(3), 39-41.

GLBA. (1999). Public Law 106-102: Gramm–Leach–Bliley Act (pp. 143). 106th Congress.

Grimaila, M. R., & Kim, I. (2001). An undergraduate business information security course and laboratory. *Journal of Information Systems Education, 13*(3), 189-195.

Houseman, K., & Nevle, M. (2009). Financial Institutions Must Rethink Their Remote Deposit Capture Support Models (pp. 1-9): First Data Corporation.

Jesus, A. D. (2006). Remote Deposit Capture – A Practical Guide for Corporations. 1-20.

Johnson, T. (2009). Remote Deposit Capture. *Credit Union Management, 32*(2), 22-25.

Joseph, C. (2011). Understanding and Addressing Risk of Merchant Capture (pp. 16-17). West Virginia.

Lennon, D. (1996). Tele-banking. *Europe*(359), 38.

Levitin, A. J. (2009). REMOTE DEPOSIT CAPTURE: A LEGAL AND TRANSACTIONAL OVERVIEW. *Banking Law Journal, 126*(2), 115-122.

MarketsandMarkets. (2010). A New Look at Merchant Remote Deposit Capture (pp. 1-82). Markets and Markets.

McLaughlin, S. (2008). How & Why to Implement a Remote Deposit Capture Program. *The RMA Journal, 90*(7), 22-24.

Meara, B. (2008). State of Remote Deposit Capture 2008: Sprint Becomes a Marathon (pp. 49): Celent.

Meara, B. (2011). State of Remote Deposit Capture 2011: Signs of a Maturing Market: Celent.

NIST. (2010). Guide for Applying the Risk Management Framework to Federal Information Systems: A Security Life Cycle Approach *Information Security* (pp. 1-93): National Institute of Science and Technology.

Patel, R. (2010). Cybercrime: How to Protect Your Financial Institution. *Michigan Banker, 22*.

Podhradsky, A., Streff, K., Pauli, J., & Engebretson, P. (2011). *Restructured Information Technology Risk Assessment Model for Small and Medium-sized Financial Institutions*. Paper presented at the Proceedings of the 11th Annual Hawaii International Conference on Business, Honolulu, HI, USA, May, 2011.

Poore, R. S. (1999). Generally Accepted System Security Principles. *Information Systems Security, 8*(3), 27-77.

Riivari, J. (2005). Mobile Banking: A powerful new marketing and CRM tool for financial services companies all over Europe. *Journal of Financial Serivces Marketing, 10*(1), 11-20.

Saleh, Z. I., Refai, H., & Mashhour, A. (2011). Proposed Framework for Security Risk Assessment. *Journal of Information Security, 2*, 85-90.

Spears, J. L., & Barki, H. (2010). USER PARTICIPATION IN INFORMATION SYSTEMS SECURITY RISK MANAGEMENT. *MIS Quarterly, 34*(3), 503-522.

Symantec. (2009). Internet Security Threat Report (Vol. XV): Symantec Corporation.

Valentine, L. (2008). The RDC experience: Captivating. *ABA Banking Journal, 100*(3), 36-40.

Appendices and Annexures

Concepts	Definitions
Confidentiality (C)	Confidentiality is the processes, policies, and controls employed to protect information against unauthorized access or use
Integrity (I)	Integrity is the processes, policies, and controls used to ensure information has not been altered in an unauthorized manner that compromise accuracy, completeness, and reliability
Availability (A)	Availability is the processes, policies, and controls used to ensure authorized users have prompt access to information, protecting against intentional or accidental attempts, to deny legitimate users access to information or systems
Volume (V)	Volume is the amount of information stored, processed, and transacted by an asset
Asset Protection Profile (APP)	Asset Protection Profile is a score calculated by adding the quantitative score for confidentiality, integrity, availability and volume for the asset
Asset Threat Score (ATS)	Asset Threat Score is a score calculated by the multiplication of Impact (I) and Likelihood (P) rating for each threat of the asset
Total Asset Threat Score (TATS)	Total Asset Threat Score is equal to the sum of all threat scores
Asset Inherent Risk Score (AIRS)	Asset Inherent Risk Score is a score calculated by taking the Total Threat Score times the Asset Protection Profile for the asset

Table 5: List of Concepts with the Definitions

Confidentiality	**High:** Information is sensitive; its disclosure would violate federal banking regulations and/or result in significant harm to the institution. **Medium:** Information is considered internal; its disclosure may violate federal banking regulations and/or result in moderate harm to the institution. **Low:** Information is for public consumption; its compromise would not be harmful to the institution.
Integrity	**High:** Accuracy of the information is critical; its modification or incorrectness would cause significant issues. **Medium:** Accuracy of the information is important, but not absolutely critical; its modification or incorrectness may cause moderate issues. **Low:** Accuracy of the information is of low concern; its modification or incorrectness may be inconvenient but could likely go unnoticed and cause few issues to the institution.
Availability	**High:** Information availability is of significant concern; recovery must be made within 24 hours. **Medium:** Information availability is of moderate concern; recovery must be made within 1 week. **Low:** Information is readily available elsewhere; recovery within 30 days is satisfactory.
Volume	**High:** There is a large amount of data regularly stored, processed, or transmitted. **Medium:** A moderate amount of information is regularly stored, processed, or transmitted. **Low:** Only a small amount of information is regularly stored, processed, or transmitted.

Table 6: CIA-V Rankings

	Bank			Merchant			TSP			
Threats	I	P	I*P	I	P	I*P	I	P	I*P	Total
Data Loss	H	M	6	M	H	6	H	H	9	21
Unauthorized Physical Access	H	M	6	H	H	9	M	M	4	19
Unauthorized System Access	H	M	6	M	H	6	H	M	6	18
Social Engineering	M	L	2	M	H	6	H	H	9	17
Malicious Software	M	M	4	H	L	3	H	H	9	16
Intentional Misuse	M	M	4	M	H	6	M	M	4	14
Eavesdropping/ Sniffing	M	M	4	M	M	4	M	M	4	12
Degraded/ Unavailable	M	M	4	L	M	2	M	M	4	10
Hardware Failure	M	M	4	L	M	2	M	M	4	10
User Error	M	M	4	M	M	4	M	L	2	10
Outsourced	H	M	6	-	-	-	-	-	-	6
Environmental Incident	M	L	2	L	L	1	M	L	2	5
Man-made/ Natural Disaster	M	L	2	L	L	1	M	L	2	5
Unauthorized Remote Access	H	L	3	-	-	-	M	L	2	5
Software Acquisition	M	L	2	-	-	-	M	L	2	4
Unauthorized Viewing	L	L	1	L	L	1	L	L	1	3
Total Asset Threat Score			**60**			**51**			**64**	**175**
Asset Inherent Risk Score			**540**			**459**			**576**	**1575**

Legend:

"I"—Impact; "P"—Likelihood; "H"—High; "M"—Medium; "L"—Low

Value:

High = 3; Medium = 2; Low = 1

Table 7: Threats of Merchant Capture Systems

	Five Controls for each Threat				
Threats	**Control 1**	**Control 2**	**Control 3**	**Control 4**	**Control 5**
Data Loss	Security Information & Event Management	Unique User Accounts	User Activity Logs	Administrator Activity Logs	Data Loss Prevention
Unauthorized Physical Access	Remote Capture User Security Controls Audit	Restricted Access Area	Surveillance Cameras	Monitored Location	Motion Detection
Unauthorized System Access	Security Information & Event Management	Unique User Accounts	User Activity Logs	Administrator Activity Logs	Firewall- Ingress Filtering
Social Engineering	Data Loss Prevention	Incident Response Program	Inactive Lockout	Website Filtering	Social Engineering Security Awareness
Malicious Software	Security Information & Event Management	Firewall- Ingress Filtering	Intrusion Detection / Prevention	Back-up Critical Data	Formal Patching Process
Intentional Misuse	Security Information & Event Management	Remote Capture User Security Controls Audit	Unique User Accounts	User Activity Logs	User Privileges & Restrictions
Eavesdropping/ Sniffing	Security Information & Event Management	Intrusion Detection / Prevention	Malware Protection	Encrypt Transmitted Data	---------------------------
Degraded/ Unavailable	Firewall- Ingress Filtering	Firewall- Egress Filtering	Incident Response Program	Back-up Critical Data	Formal Patching Process
Hardware Failure	Back-up Critical Data	Backup Recovery Test	Hardware Health Monitor	Power Conditioning	RAID
User Error	Security Information & Event Management	User Activity Logs	User Privileges & Restrictions	Administrator Activity Logs	Data Loss Prevention
Outsourced	Business Continuity Plan	Formal Third Party Selection	Formal Third Party Review	Business Continuity Plan Test	Escrow
Environmental Incident	Secure Equipment & Capable Placement	Temperature Control	Humidity Control	Environment Monitor	Food & Liquid Filtering
Man-made/ Natural Disaster	Back-up Critical Data	Backup Recovery Test	Business Continuity Plan	Business Continuity Plan Test	Redundancy/Contingency Agreement
Unauthorized Remote Access	Security Information & Event Management	Remote Capture User Security Controls Audit	Unique User Accounts	User Activity Logs	Firewall- Ingress Filtering
Software Acquisition	Business Continuity Plan	Formal Third Party Selection	Formal Third Party Review	Business Continuity Plan Test	Escrow
Unauthorized Viewing	Remote Capture User Security Controls Audit	Inactive Lockout	Monitor Placement	Privacy Filter	Clear Screen Awareness

Table 8: Top Five Controls for Threats in Merchant Capture Systems

User Questionnaire: Impacts and Likelihood of Threats

The questionnaire requests you to evaluate the likelihood and impacts of each threat provided to you by our Expert Group arranged in alphabetical order. There are no right/wrong answers. It is very important that honest evaluations are indicated.

Please choose “H”, “M” or “L” for High, Medium, and Low respectively for the likelihood (P) and impact (I) for the following identified threats for three actors: bank, merchant and technology service provider as third party.

	Bank		Merchant		Third Party	
Threats	Impact (I)	Likelihood (P)	Impact (I)	Likelihood (P)	Impact (I)	Likelihood (P)
Data Loss	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L
Degraded / Unavailable	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L
Eavesdropping / Sniffing	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L
Environmental Incident	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L
Hardware Failure	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L
Intentional Misuse	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L
Malicious Software	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L
Man-made / Natural Disaster	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L
Outsourced	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L
Social Engineering	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L
Software Acquisition	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L
Unauthorized Physical Access	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L
Unauthorized Remote Access	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L

Unauthorized System Access	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L
Unauthorized Viewing	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L
User Error	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L	☐ H ☐ M ☐ L