Economic Perspectives

2 Economic Perspectives special issue on payments fraud: An introduction

7 Payments Fraud: Perception Versus Reality—A conference summary

17 Fraud containment

22 Data security, privacy, and identity theft: The economics behind the policy debates

31 Perspectives on retail payments fraud

37 Divided we fall: Fighting payments fraud together

43 An examination of the fraud liability shift in consumer card-based payment systems

50 Vulnerabilities in first-generation RFID-enabled credit cards
<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td><em>Economic Perspectives</em> special issue on payments fraud: An introduction</td>
<td>Gene Amromin and Richard D. Porter</td>
</tr>
<tr>
<td></td>
<td></td>
<td>This article provides an overview of this special issue of <em>Economic Perspectives</em>, which presents selected papers based on the proceedings of the Federal Reserve Bank of Chicago’s eighth annual Payments Conference, Payments Fraud: Perception Versus Reality, held on June 5–6, 2008.</td>
</tr>
<tr>
<td>7</td>
<td>Payments Fraud: Perception Versus Reality—A conference summary</td>
<td>Tiffany Gates and Katy Jacob</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The authors highlight key issues from the presentations, keynote addresses, and open floor discussions at the Federal Reserve Bank of Chicago’s eighth annual Payments Conference.</td>
</tr>
<tr>
<td>14</td>
<td>Agenda: 2008 Payments Conference</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Payments Fraud: Perception Versus Reality</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Announcement: 2009 Payments Conference</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Fraud containment</td>
<td>Bruce J. Summers</td>
</tr>
<tr>
<td>22</td>
<td>Data security, privacy, and identity theft: The economics behind the policy debates</td>
<td>William Roberds and Stacey L. Schreft</td>
</tr>
<tr>
<td>31</td>
<td>Perspectives on retail payments fraud</td>
<td>Steve Malphrus</td>
</tr>
<tr>
<td>37</td>
<td>Divided we fall: Fighting payments fraud together</td>
<td>Mark N. Greene</td>
</tr>
<tr>
<td>43</td>
<td>An examination of the fraud liability shift in consumer card-based payment systems</td>
<td>Duncan B. Douglass</td>
</tr>
<tr>
<td>50</td>
<td>Vulnerabilities in first-generation RFID-enabled credit cards</td>
<td>Thomas S. Heydt-Benjamin, Daniel V. Bailey, Kevin Fu, Ari Juels, and Tom O’Hare</td>
</tr>
</tbody>
</table>
**Economic Perspectives special issue on payments fraud: An introduction**

Gene Amromin and Richard D. Porter

In this special issue of *Economic Perspectives*, we present selected papers based on our recent conference, Payments Fraud: Perception Versus Reality, hosted by the Federal Reserve Bank of Chicago on June 5–6, 2008. The conference brought together decision-makers from the banking, payments, legal, regulatory, and merchant communities for a wide-ranging discussion of the threats to the security of the payments system and how those threats might best be addressed.

The volume starts with an extensive summary of conference presentations, keynote addresses, and open floor discussions, written by Tiffany Gates and Katy Jacob. In order to give a sense of the intense back-and-forth exchanges that took place during this day-and-a-half-long event, the authors structure their summary around the broad themes of the discussion rather than simply presenting a chronological account. The themes are as follows: organizational structures for management of fraud risks; technological innovation; alignment of incentives for fraud prevention among consumers, merchants, and payments providers; and regulatory policies.

Gates and Jacob’s article highlights the challenges involved in bringing the various constituencies together to forge common ways to address fraud in payments systems. Gates and Jacob find that payments fraud cannot be eliminated without decreasing the openness and efficiency of the payments system. In the current environment, technological innovations have enabled system participants to enhance payments security, at the same time that technology has made it easier for criminals to perpetrate payments fraud remotely. Practitioners are constantly weighing the costs and benefits of payments fraud mitigation and are looking to the public sector to offer guidance and support. As the industry combats payments fraud, companies are banding together to find common solutions. For instance, throughout the conference, financial industry participants emphasized the concept of enterprise-wide fraud management, while many also acknowledged the difficulties faced by small merchants and many financial institutions in fashioning such holistic strategies. A number of legal professionals stressed the detrimental effects of legacy laws and regulations that evolved independently around individual payment product lines. Together, these viewpoints contributed to a budding consensus on the importance of dedicated high-level executive involvement in payments fraud management and of outsourcing development of fraud prevention tools to specialized entities.

The rest of this volume is devoted to articles that address in greater detail some of the key topics discussed at the conference. The contributors of these papers span the spectrum of thought leaders in combating payments fraud—industry experts in fraud detection systems, legal professionals, academic researchers in economics and technology, and senior officials of the Federal Reserve System.

The first article is written by Bruce J. Summers. His paper provides a synthesis of the approaches of practitioners and economists to thinking about the problems in containing retail payments fraud. As Summers makes clear, these approaches differ somewhat for reasons that have to do with both perspective and analytical framework. Yet, both parties are integral in formulating a coherent public policy response to the problem of payments fraud.

In particular, payments industry practitioners tend to regard fraud as a persistent but manageable problem that requires both unrelenting attention and significant expenditures. These expenditures on fraud mitigation...

---

Gene Amromin is a senior financial economist and Richard D. Porter is a vice president, senior policy advisor, and the director of the payments team in the Financial Markets Group at the Federal Reserve Bank of Chicago.
have resulted in declining rates of fraud losses. Still, there is concern that maintaining such results in the future will require ever-expanding expenditures. Part of this argument rests on the view that fraud threats to electronic payments networks arise globally, are increasingly sophisticated, and propagate quickly. The bottom line is that for the practitioners, payments fraud is part of the cost of doing business, which can be ameliorated by pooling fraud prevention efforts. To underscore this view, Summers reports consensus recommendations from a recent industry roundtable (which is a focus of the article by Malphrus later in this volume). Those recommendations call for information sharing, better authentication technologies, and adoption of standards—ideas that take into account the economies of scale and scope in fraud prevention, though not necessarily the conflicting incentives among the many participants in payments networks.

Economists tend to think of the payments system as a vast network of participants whose divergent incentives generate considerable spillovers from their own actions. The effects of these spillovers are frequently not fully appreciated by the participants because market prices fail to convey the extent of the spillovers to them. An example of such an “externality” is when a consumer (or merchant) fails to be appropriately vigilant about data security because limited liability rules (or existing penalties for breaching network security protocols) do not signal the extent to which such actions affect other participants in the payments system. Moreover, because fraud prevention by one party usually improves the experience of everyone else, there is ample incentive for other participants to “freeload.” As a result, there may be drastic underprovision of fraud prevention in the aggregate. This framework forces economists to take a systemic view of payments networks, focusing on ways in which policy can better align the incentives of all participants. One important implication of this view is that incentives (or regulation) must be appropriately allocated across the payments network because the entire system is only as strong as its “weakest link.”

The juxtaposition of the views of practitioners and academic economists gives rise to an interesting and provocative observation that industry practitioners’ relatively sanguine view of risk is partly attributed to their focus on practices of their own firms. Thought leaders in the industry are keenly aware that the interconnectedness between the many players in the payments space poses risks that are not directly observable by any individual participant. Yet, their primary responsibility for managing fraud at their enterprises may instill a somewhat false sense of security. This comparison also points to a potentially key role of the Federal Reserve in bridging the gap, since it is both a research center and a major payments provider.

The next three articles in this volume are, in some sense, elaborations on Summers’ conclusion. The first of these, by William Roberds and Stacey L. Schreft, lays out an economic framework for thinking about fraud in payments networks. The second, by Steve Malphrus, gives the industry view on the current state of efforts in retail fraud management. The third, by Mark N. Greene, addresses the nature of payments fraud and the need for coordinated efforts in fighting it.

Roberds and Schreft start out by noting the inevitable trade-off between more efficient payments markets and loss of privacy. On the one hand, an economy grows, paying for transactions in cash becomes prohibitively expensive and inefficient. On the other hand, credit- and debit-based transactions between parties that typically do not know each other are impossible without exchanging some information that verifies both the identity and the creditworthiness of the parties. The resulting transfer of information back and forth presents opportunities for fraud. Still, without such transmission of private information modern payments systems would be infeasible, thereby forcing payments activity onto the slower rails of cash-facilitated exchanges of yesteryear.

What is the proper balance then? As Roberds and Schreft argue, it is useful to think of this balance as “confidentiality” of payments transactions. “Confidentiality” thus consists of “data informativeness” (how much identifying information is exchanged between parties) and “data integrity” (how well this information is protected). If you tilt the balance too much toward safeguarding privacy (that is, lessen data informativeness), then the wheels of credit-based and remote transactions grind to a halt. But if you tilt it too much toward being absolutely certain about the identity and creditworthiness of the consumer, then you may transmit so much data that you increase the incidence of and losses from fraud.

To get closer to the answer, the authors lean on economic theory. They note that neither information nor integrity diminishes with repeated use. For example, say Wal-Mart knows Mr. A is not a fraudulent actor and, therefore, processes payments it receives from him. The fact that Wal-Mart has this information does not diminish the value of the information to Home Depot. These seemingly abstract concepts matter because they help us think about the way in which “confidentiality” can be provided efficiently. In particular, the fact that the two attributes of “confidentiality” do not wear down with consequent use implies that, at some point, the marginal cost of providing it is close to zero.
This insight implies that optimally there will be only a few large producers of “confidentiality” that are able to leverage vast economies of scale in building networks for collecting and transmitting necessary information securely (for example, credit bureaus, card networks, and so on). However, since information must be exchanged among many different parties (for example, merchants, issuers, and consumers) in order to have value, the potential for conflicts of interest is large. Some parties may have weaker incentives to safeguard information and thus become the “weak link” that compromises the entire system. Some parties may choose to freeload on data integrity efforts of others because the cost of fraud or loss of “confidentiality” is not proportionately allocated.

The authors’ analysis of “confidentiality” further points to the proper role for policymakers in fostering efficient (but not fraud-free) payments systems. Public policy should aim to resolve the potential for conflicts of interest through coordination, judicious imposition of standards, and proper allocation of legal incentives (as discussed in greater detail in Douglass’s article later in this volume). The public sector should not focus on duplicating the job done by the private sector in collecting, verifying, and processing payments-related information. As Roberds and Schreft underscore, the ultimate goal of regulators should be to strike the proper balance between privacy and efficiency.

Malphrus’s article summarizes the results from a special roundtable discussion on retail payments fraud, which was held at the Federal Reserve Bank of Minneapolis in March 2007. The participants reported that, despite the declining use of checks, check payments still generated the largest number of fraud attempts. They emphasized that criminals are continually searching for weaknesses in fraud detection and prevention practices. Many thought that banks and businesses needed to adopt a holistic approach to detecting and preventing retail payments fraud, being ever mindful of the overall fraud landscape across all of their operations. The roundtable participants shared a number of suggestions for improving the industry’s ability to detect and prevent retail payments fraud, including better protection of customers’ personal and financial data. They recommended more effective sharing of best practices with respect to fraud detection and prevention within the industry. Industry leaders specifically discussed the effectiveness of PIN (personal identification number) and chip technology. Some stated that fraud rates on PIN debit cards are significantly lower than those for other payment types, and advocated a more widespread application of PIN security to card payments.

Malphrus also shares some thoughts on new account fraud that has featured prominently in recent discussions on retail payments systems’ vulnerabilities. Allowing customers to open accounts electronically, as opposed to in person at a bank, clearly offers the potential for fraud. However, this risk can be mitigated by making use of various technologies; for example, software can identify the geographical location of the user’s computer, and device identification tests can be subjected to further fraud screening. All of these technologies are currently operational, and their widespread adoption is likely to make a considerable difference in mitigating a particular aspect of retail payments fraud.

Malphrus also highlights an increasingly important aspect of policymaking—the need to protect privacy while countering terrorist financing and money laundering. As different agencies (for example, the Central Intelligence Agency and the Federal Reserve) cooperate to combat these threats, they must be vigilant about how they exchange information about U.S. citizens. Moreover, as those perpetrating illicit activities increasingly attempt to leverage existing payments networks, the need for cooperation between the private sector and government agencies becomes all the more important.

The next contributor, Greene, represents Fair Isaac Corporation—a leading provider of automated identification procedures for prospective fraud on electronic payments networks. Greene’s role gives him a clear perspective on the nature of the current threats to the payments system. As Greene argues in his article, greater cooperation among the various payments system participants is necessary to combat fraud. Increasingly, modern day fraudsters operate globally, often outside the jurisdiction of the U.S., and they are well organized and well financed.

In particular, Greene warns that adopting piecemeal solutions that focus on individual payments segments or regions would be inadequate to beat the scams. While payments providers may have an incentive to differentiate their products and seek competitive advantage, they need to find ways to cooperate with each other in sharing information or developing standards that would help lessen the problem. He suggests that piecemeal solutions are like pushing on a balloon—they may impede fraud in the particular targeted segment or region but quite often at the expense of increasing fraud elsewhere.

Instead, Greene advocates a fraud protection system that works like a burglar alarm, covering all the openings—“doors and windows”—since fraudsters will always make the most of the weakest link. He argues it is possible to build better models of fraud.
containment that profile not only individuals but also devices and merchants. With this platform, one could successfully identify uncharacteristic and possibly fraudulent payment behavior. Such modeling exercises would be more effective if they could be “trained” (that is, estimated on large amounts of current real-world data); this might require the cooperation of various payments system participants. Some public sector cooperation might be desirable to remove the concern that such industry data-sharing exercises constituted collusive practices.

In the question-and-answer session that followed Greene’s keynote address at the conference, Greene raised the possibility of a mass compromise to the payments system by fraudsters. In a typical card compromise where the information could be stolen for, say, 25,000 cards, Greene acknowledged there is often only a limited amount of resulting damage—perhaps on the order of 400 fraudulent transactions. He suggested that the outcome could be considerably larger, with perhaps as many as 4 million fraudulent transactions generated on the same (25,000) card base. In this circumstance, Greene argued the systemic risk to payments could be huge. Moreover, he stressed that the industry is not prepared to deal with such a contingency. This might require a joint public–private initiative to scope out the problem and propose solutions.

The next article in this volume, by Duncan B. Douglass, takes us back to the central role that proper allocation of incentives plays in the efficient functioning of payments systems. Douglass focuses on ways in which the current framework of public laws and private network rules distributes fraud liability among the three principal sets of participants—consumers, merchants, and card issuers. Although the discussion centers on signature-based credit and debit cards, its implications are readily extended to other payments instruments.

The public law framework that governs the legal liability for fraud losses is based on the Truth in Lending Act (TILA) and the Electronic Fund Transfer Act (EFTA), as well as the associated Federal Reserve Board Regulations Z and E. As Douglass points out, the primary goal of these laws and rules is to effectively absolve the consumer from liability for losses related to fraud, regardless of whether a consumer’s own behavior contributed to fraud in the first place. Although a lack of care on the part of consumers often contributes to fraud, Douglass argues that making consumers bear more of the consequences for their actions is not realistic. This owes both to the political environment and, more importantly, to the desire to instill confidence in the security of card transactions among consumers.

The private card networks’ rules take over from where public laws stop—by setting cardholder liability to zero—and proceed to further allocate fraud loss liability between merchants and card issuers. In brief, the rules effectively assign liability for losses in card-present transactions to issuers and in card-not-present transactions to merchants. Douglass emphasizes that this joint framework of public laws and private rules leads to several predictable outcomes that make system-wide fraud prevention efforts somewhat inefficient. To paraphrase, consumers never care too much about safeguarding their transactions, merchants try to exercise due diligence primarily in card-not-present environments, and card issuers are concerned mostly with point-of-sale transactions. Each party thus has ample incentives to undermine the efforts of the other—for example, merchants not verifying signatures at the point of sale.

Douglass illustrates this dynamic with the example of the failed adoption of networks’ payer authentication programs. Although these programs are effective in reducing online fraud, consumers, who bear no responsibility for fraud, have balked at merchants’ efforts to adopt these measures. For their part, card issuers, which bear less of a burden for fraud in card-not-present transactions, were content to sit on the sidelines and not force their customers to enroll in such programs.

The final article in the volume, by Kevin Fu, Thomas S. Heydt-Benjamin, Daniel V. Bailey, Ari Juels, and Tom O’Hare, focuses on technological vulnerabilities in a newly popular set of payment instruments—devices that use RFID (radio frequency identification), such as credit cards. Such cards offer the promise of speedier contactless transactions at the checkout or gas station and, unlike traditional magnetic stripe cards, require only physical proximity between the card and the associated reader.

Fu and his co-authors demonstrate that the more convenient retail experience provided by RFID devices over magnetic stripe cards may come at the price of several vulnerabilities in RFID’s first-generation incarnations. Using their toolkit as electrical engineers, the authors find that all of the 20 million RFID-enabled cards currently in circulation are subject to privacy invasion. The cards can be scanned and private information can be removed by the fraudsters without the awareness or consent of the cardholders.

These vulnerabilities should not necessarily be viewed as a fatal indictment of the technology; rather, they represent what might be expected for a work in progress. If successful, RFID technology will overcome these vulnerabilities along its developmental path.
New (and ultimately successful) payments innovations do not necessarily provide full fraud protection capabilities at launch but often gain them over time as they scale up efficiently. The history of PayPal illustrates this point quite nicely.¹ We hope that the message delivered by Fu and his co-authors will rouse the card manufacturers to address these challenges quickly.

Each of the articles collected in this volume offers a specific insight into the current state of efforts in combating retail payments fraud. The articles also outline a number of ways in which these efforts can be made more successful at a systemwide level and offer a methodological framework for thinking about the problem. We hope this work will provide a valuable basis for ongoing discussions on how we can develop and coordinate public and private responses to the pressing need to manage payments fraud risk.

NOTES

Payments Fraud: Perception Versus Reality—
A conference summary

Tiffany Gates and Katy Jacob

An overview of payments fraud

Payments fraud can be broadly defined as any activity that uses information from any type of payments transaction for unlawful gain. Such fraud can be perpetrated on any type of payments device, including credit and debit cards, cash, checks, online or mobile payments, and automated clearinghouse (ACH) transactions. Payments fraud can be committed knowingly by a consumer (first-party fraud), or consumers can be victimized by fraudsters operating within financial institutions or as part of criminal enterprises (third-party fraud). Payments fraud has received extensive attention in the popular press and in public policy venues recently, and the payments industry is fighting the perception that fraud is now occurring at unmanageable levels. While there has been increasing emphasis on all types of payments fraud, fraud perpetrated by criminals has received special attention of late.¹

Fraud is a very real threat to the payments system’s efficiency. According to one recent report, 71 percent of surveyed organizations experienced payments fraud in 2007, and over one-third of those firms reported financial losses stemming from the fraudulent activity.² As another example of the size of the payments fraud problem, in a 2007 data breach involving TJX Companies Inc. (the holding company of retailers T. J. Maxx, Marshalls, Winners, HomeGoods, TK Maxx, A. J. Wright, and HomeSense), 45,700,000 credit card and debit card account numbers were stolen, along with 455,000 merchandise return records containing customer names and driver’s license numbers. Latest reports allege that an additional 48 million people have been affected for a total of over 30 percent of the entire U.S. population. The situation has cost TJX Companies Inc. more than $130 million in settlement claims. The breach was a worldwide effort perpetrated by criminals from the United States, Eastern Europe, and China. The U.S. Department of Justice has arrested 11 people in this case, which is the largest hacking and identity theft case ever prosecuted by the department.³

As more payments become electronic, the size and scope of payments fraud has grown, in part because the relevant parties in a payments transaction do not know one another. Information about those parties is vital to prevent fraud and enable legitimate transactions. However, as innovations in payments technology have made authentication of information more reliable, other technological innovations have made that information more widely available and subject to abuse. Fraud such as counterfeiting or check forgery has always had a global reach. However, payments fraud used to be much more reliant on physical connections between parties, such as the theft of an individual checkbook or credit card.

Today, modern databases, online information sharing, and increased access points have opened up opportunities for sophisticated criminal gangs to perpetrate fraud from remote corners of the globe. Further, the growing presence of nonbanks and third-party service providers means that regulated financial institutions must consider the security of those providers. At the same time, new laws and standards are being developed for payment activities and instruments. While the continual refining of systems and rules arguably makes payments easier and more efficient, the fast pace of change can compound fraud potential as fraudsters hunt to exploit the weakest link in the emerging systems.

In this complex environment, market participants and governments must determine whether new payment

Tiffany Gates is a supervision analyst in the Banking Supervision and Regulation Department at the Federal Reserve Bank of Chicago. Katy Jacob is a research specialist in the Financial Markets Group at the Federal Reserve Bank of Chicago. The authors thank the Chicago Fed’s payments team for their help in producing this article.
types carry excessive fraud risk; who is liable when payments fraud occurs; how losses are allocated; what consumer protections should be in place; how notification of fraud should be handled; and how standards should be defined to minimize the incidence of fraud. It is a tall order, but payments providers must also identify consumers whom they have never met and authorize electronic transactions from which they might be far removed. And, increasingly, they must do this in real time.

To explore the problem of payments fraud, the Federal Reserve Bank of Chicago organized its eighth annual Payments Conference around the topic. The conference, Payments Fraud: Perception Versus Reality, took place on June 5–6, 2008. In this article, we summarize the conference and focus on the following themes: why the industry is worried about payments fraud; managing fraud risks; the impact of technology and innovation on fraud; responsibilities and incentives for fraud prevention; and public sector involvement in mitigating payments fraud. We note that market participants agree that payments fraud cannot be eliminated without risking the viability of certain payment channels, but also find that close industry collaboration, properly aligned incentives, technological innovations, and active risk management can lessen fraud’s ill effects.

Why worry about payments fraud?

Fraud degrades operational performance and increases cost—not only for the parties to the transactions whose payments are disrupted, but also for the payments system as a whole. Indeed, payments networks are vulnerable to fraud at any point in a payments chain, and fraudsters often attempt to exploit the weakest link in that chain. One of the foremost concerns is the potential for a single data breach or compromise to disrupt an entire payments system. According to conference panelist Jeff Schmidt, an independent consultant, it is possible for a single data breach to affect multiple layers in the payments system and disrupt the efficient operation of the entire system if confidence in the system is lost.

Further, Mark Greene, Fair Isaac Corporation, raised the possibility of a mass compromise of significant components of the U.S. payments industry. Greene said that the industry is not prepared for a mass attack wherein fraudsters target multiple companies simultaneously through hacking and sophisticated phishing techniques. These threats have the potential not only to harm a financial institution but also to degrade the payments system globally. Bruce Summers, a payments system and technology management consultant, questioned whether the marketplace alone could contain fraud and protect the payments system as a whole if such a mass compromise were to occur. Indeed, Allison Edwards, Fiserv EFT, commented that the payments industry was completely caught off guard by the aforementioned 2007 TJX Companies data breach because of its size and scope.

It is important to note that there is a distinction in the payments industry between actual fraud that has been perpetrated and potential fraud from compromised information that might not necessarily result in criminal activity. Ellen Richey, Visa Inc., claimed that the number of compromise incidents in the United States is rising, while other analysts contend that only the reporting of these incidents is increasing. Regardless of the magnitude of growth, industry leaders are concerned about both stopping compromises from occurring and ensuring that significant fraud does not take place when compromises do occur. Conference panelists maintained that when such fraud happens, consumer confidence can only be restored by a fast and thorough industry response.

Managing fraud risks

As it stands, many in the industry find it difficult to gauge the full impact of fraud on the payments system. Richey applauded the payments industry for doing a good job in stemming the tide of increasing fraud attacks, stating that global fraud rates in the card industry have remained largely constant since 2002. Others at the conference argued that, while the total amount of fraud has gone down, the impact of the fraud that does occur has become more costly to society. Summers commented that many in the payments industry argue that today’s level of fraud protection is sufficient, and noted that few market participants seem dissatisfied with the overall state of payments fraud. Some players view fraud as just another cost of doing business, though according to several conference participants, that view is being overshadowed by an urgent need to keep fraud under control.

According to David Poe, of Edgar, Dunn, and Company, many payments participants often make suboptimal risk-management business decisions because the true cost of fraud is misunderstood. Most analysts only take account of fraud losses to issuers (banks that issue payment cards to consumers or businesses) when tallying fraud costs. Poe noted that the monthly benchmarks for issuers’ fraud losses are approximately 0.07–0.08 percent of transaction volumes. Fraud losses to acquirers (banks that process card payments for merchants) from chargebacks are also of about the same magnitude. Poe echoed Greene by noting that statistics on issuers’ credit card losses from first-party fraud showed that fraud could account for as much as 10 percent of their credit losses if correctly categorized.
Moreover, opportunity cost—where consumers pass up one payment option or company in favor of another because of perceived security concerns—is arguably the biggest cost of fraud and the most difficult to quantify. It is the largest potential risk in that customers might not use a payment product at all, or might not use the product in the appropriate way, because they do not trust that the payment instrument is secure.

When determining the true cost of payments fraud, analysts sometimes also fail to count the cost borne by issuers, acquirers, and merchants to manage fraud risks. Bob Ledig, of Fried, Frank, Harris, Shriver, and Jacobson LLP, stated that the costs of fraud cannot be limited to direct costs borne by any one party in the payments system. Rather, resource, compliance, enforcement, reputation, and litigation costs must also be taken into consideration. He noted that data security should be an inherent part of the payments vehicle, rather than a separate line of business. These comments about the true price of payments fraud raise the possibility that there may be some type of market failure in the payments system, wherein the nature of fraud is so complex that firms are unable to price it correctly.

To keep costs down and to better manage the risk associated with payment channels and instruments, financial institutions are looking to incorporate an enterprise-wide approach to fraud management. Challenges arise because lines of business have historically been developed as independent silos. Judith Rinearson, of Bryan Cave LLP, stressed that payments laws and regulations have largely emerged around individual product lines, making it difficult to implement enterprise-wide solutions. Many audience members commented that small merchants also struggle to implement enterprise-wide solutions, as they lack the resources to obtain high-end fraud prevention tools. The transition to an enterprise-wide approach to fraud mitigation is driven by governance and culture. Conference participants felt that the comparative handful of organizations that have appointed “payment czars” have been more effective in looking at payments fraud across the institution as a whole. Yet, if an institution has a deeply siloed governance and organizational structure, it is difficult to develop consistent, cost-efficient business processes across different product lines.

Greene urged the industry to take note of the “balloon effect” in payments fraud. Namely, once fraud begins to decrease in one payment method, criminals often shift focus to another part of the payments system, where fraud rates begin to rise. Audience members commented that fraud might also shift among regions or nations. Some speculated that the increasing use of chip and PIN (personal identification number) technology in Europe and Canada might lead criminals to focus on countries that rely more heavily on older magnetic stripe technology, such as the United States. These different types of fraud shifts could lead to misperceptions about what is truly occurring in the system as a whole, and they are especially important to consider when new payments technologies enter the market.

Payments technology and innovation

On the one hand, technological innovations have enabled market participants to authenticate payments information more accurately in real time, greatly enhancing the security of electronic payments transactions. On the other hand, the speed of payments innovation can accelerate fraud risks. Traditionally, the payments industry has been slow to manage technology, while fraudsters have quickly adapted to the new channels available. Poe reinforced the idea that technology has made fraud easier to commit on a wide scale, citing the increases in phishing, pharming, skimming, and other fraud tactics that often rely on remote or card-not-present transactions.7

According to Kevin Fu, University of Massachusetts Amherst, phishing is one of the biggest security problems on the Internet. It is certainly the easiest way a spammer (one who uses electronic messaging systems to indiscriminately send unsolicited bulk messages) can infiltrate thousands or millions of compromised machines around the world. If just a tiny fraction of the people spammed respond, the spammer can obtain quite a bit of sensitive information that can be used to perpetrate fraud. Richey went further by saying that the top vulnerabilities in the payments system are the storing of prohibited data; out-of-date systems; perimeter security; weak wireless security systems; and structured query language (SQL) injection attacks.8 These vulnerabilities can only be addressed if every participant in the payments system is accountable and vigilant about protecting data, upgrading systems, and monitoring its own staff and its partner firms. However, upgrading software and infrastructure can be quite costly. In some cases, technology enhancements happen so quickly that companies, especially small merchants and processors, have little time to react.

Consumer perceptions of fraud risks can also directly impact the success of a new payment method. Greene noted that consumers’ perception that mobile and contactless payments are more prone to fraud has apparently stunted the growth of those payment channels in the United States. Mobile payments are payments that are initiated by a mobile device, such as a mobile phone.9 A contactless payment device, such
as a card or fob, uses radio frequency identification (RFID) or near field communication (NFC) technology to make secure payments. The embedded chip and antenna enable consumers to wave their payment device over a reader at the point of sale. Both RFID and NFC payment methods are relatively new in the U.S. market, and it should be noted that it often takes time for consumers to adopt any new instrument or market. Bruce Cundiff, Javelin Strategy and Research, echoed the sentiment that risk adversely affects consumer adoption of these new payment instruments. Because repairing the damage done by payments fraud is becoming more complex for consumers, many are reluctant to switch to a new payment method. For example, in a recent Javelin survey, 65 percent of those who said they did not want to use contactless cards named security fears as the number one reason, and 33 percent of those surveyed viewed mobile banking as too risky.10

Cundiff pointed out a marked change in the way that consumers perceive the security efforts of their financial institutions. Consumers now want to be more engaged in security measures and view companies that allow them to be engaged through account alerts or verification calls as being more reliable. Rinearson agreed, arguing that many consumers are confused about fraud prevention features of different payment cards, such as prepaid cards versus debit or credit cards. For example, consumers might find out about fraudulent transactions from billing statements for their debit cards or credit cards, but would not have such information for a number of prepaid cards.

Payments fraud can affect the availability of new products as well. Payments providers might be hesitant to innovate in an area where unknown fraud risks exist. Paul Tomasofsky, Two Sparrows Consulting LLC, said that the newly emerging decoupled debit field faces challenges as issuers work out several potential risks. A decoupled debit card is a debit card issued by a nonbank or bank that is linked to a demand deposit account that the issuer does not own. The payments are processed on the automated clearinghouse network, which mimic minicomputers and store enormous amounts of data, can mitigate the security risk of fraud perpetrators regularly exploit new technologies to their benefit, but payments providers are working to find ways to exploit technology for fraud resolution as well. These firms are incorporating technology into the broader design of their fraud detection mechanisms. Edwards noted that “neural” networks are helping companies to manage their risk profiles more conservatively by adding the elements of time control and customer targeting. Fu discussed the ways that RFID technology in contactless cards and mobile payment devices can allow for sophisticated tracking in order to reduce fraudulent transactions. The RFID tags, which mimic minicomputers and store enormous amounts of data, can mitigate the security risk of handing over your card to someone who may want to compromise the information contained on it.

Cundiff pointed out a marked change in the way that consumers perceive the security efforts of their financial institutions. Consumers now want to be more engaged in security measures and view companies that allow them to be engaged through account alerts or verification calls as being more reliable. Rinearson agreed, arguing that many consumers are confused about fraud prevention features of different payment cards, such as prepaid cards versus debit or credit cards. For example, consumers might find out about fraudulent transactions from billing statements for their debit cards or credit cards, but would not have such information for a number of prepaid cards.

Payments fraud can affect the availability of new products as well. Payments providers might be hesitant to innovate in an area where unknown fraud risks exist. Paul Tomasofsky, Two Sparrows Consulting LLC, said that the newly emerging decoupled debit field faces challenges as issuers work out several potential risks. A decoupled debit card is a debit card issued by a nonbank or bank that is linked to a demand deposit account that the issuer does not own. The payments are processed on the automated clearinghouse network, which mimic minicomputers and store enormous amounts of data, can mitigate the security risk of fraud perpetrators regularly exploit new technologies to their benefit, but payments providers are working to find ways to exploit technology for fraud resolution as well. These firms are incorporating technology into the broader design of their fraud detection mechanisms. Edwards noted that “neural” networks are helping companies to manage their risk profiles more conservatively by adding the elements of time control and customer targeting. Fu discussed the ways that RFID technology in contactless cards and mobile payment devices can allow for sophisticated tracking in order to reduce fraudulent transactions. The RFID tags, which mimic minicomputers and store enormous amounts of data, can mitigate the security risk of handing over your card to someone who may want to compromise the information contained on it.
Greene mentioned the rise of profiling mechanisms that compile fraud patterns for specific merchants as well as in geographically dispersed payment devices and terminals. These mechanisms can be used in adaptive models that keep up with changes in fraud patterns; they allow users to dynamically change model weights to suit their needs. He argued that fraud prevention should not be viewed as providing a competitive advantage for any firm. Otherwise, fraud becomes too great of a collective problem. Fu also supported the use of open source RFID technology rather than the proprietary systems that companies are now pursuing. This idea furthers the notion that collaboration is required to combat fraud in the payments system.

Responsibilities and incentives for fraud prevention

Conference participants noted that, as consumers, merchants, and payments providers struggle with the issue of payments fraud, the goal is not to eliminate fraud but rather to generate better risk-management practices that strike a balance between allowing for risks in the payments system and dictating payments choices. Speakers at the conference were unanimous in the view that collaboration within and among companies is a necessary aspect of successful payments fraud mitigation. Security is expensive to achieve and maintain. Therefore, it can result in indirect but nonetheless real costs to consumers if those costs are transferred. Cooperation is thus not only desirable but also necessary.

According to the conference speakers, in order to be effective, payments fraud mitigation efforts must recognize the need to include all members of the system. To do this, incentives must be properly aligned. Market participants must have sufficient reasons to care about fraud mitigation. For instance, Mallory Duncan, National Retail Federation, argued that we currently have pricing and protection scenarios that encourage customers to use signature-based payment cards rather than PIN-based cards, leading to perverse incentives to use a payment vehicle that is perceived to be less secure. Moreover, banks and merchants often base their preference for different payments mechanisms on narrow cost reasons, thereby overlooking the hidden costs embedded on the security side.

Duncan also noted that if merchants do not feel that they are directly benefitting from increased data security, they will not be willing to pay for new security infrastructure. He said that it is very difficult for merchants to keep up with constantly changing payments rules, as merchants are being asked to handle payments technologies that are outside of their core competencies.

Schmidt countered that today all industries face security issues and that compliance is not specific to payments.

Several conference participants suggested that one solution to the problem of data storage standards is to be parsimonious with payments data and store only as little as the law requires. Mark Michelon, Orbitz Worldwide, explained that fraud detection needs to be automated in order for merchants to do it in a cost-effective manner. Richey elaborated by stating that effective authentication can make stolen data useless. Schmidt agreed, noting that there is so much payments data available that fraud solutions should not focus on limiting data but rather on making the data less meaningful. Public disclosure of sensitive data devalues the data for fraudsters and essentially halts the fraud. In other words, if data such as Social Security numbers are not deemed to be highly confidential, the impact of having such data stolen will not be as great. Alternative types of data include addresses or zip codes; according to Richey, these are quite effective authentication tools in many instances.

Schmidt suggested that incentives for fraud prevention should be aligned with responsibility and that potential victims should be given good reasons to care about protecting their own payments data. Several presenters commented on consumers’ relative lack of incentives in preventing payments fraud, especially in the credit card market where zero liability policies protect consumers from virtually all losses. Duncan Douglass, of Alston and Bird LLP, argued that there needs to be a realistic price tag placed on risk. Currently, he said, attorneys work with payments systems to help them decide if paying to eliminate risk is worth the cost. Payment channels rely on customer confidence for survival, but there is a moral hazard problem when customers have little incentive to be careful with data. Michelon stated that one solution to this problem is consumer education about payments fraud and data protection. While these efforts can be useful, in order for them to have meaningful effects, all actors in the payments system must have similar incentives to avoid payments fraud.

Indeed, if fraudsters are to stay in business, it would seem to be in their best interest to avoid creating a situation where a mass compromise would disrupt the payments system as a whole or destroy a specific payment channel that had previously proven lucrative for them. Marsha McClellan, United States Attorney’s Office for the Northern District of Illinois, remarked that there should be real consequences for committing payments fraud that are significant enough to make criminals think twice. She stated that it is difficult to prosecute a payments fraud case because
of the electronic nature of the crime, which usually means there is not much physical evidence. Moreover, many consumers have a hard time pinpointing compromised information. McClellan suggested that monetary incentives were the most likely way to deter fraud. United States Attorneys have the authority to seize the proceeds of criminal activity even before prosecutions occur. If funds are seized, criminals lose the ability to continue their operations. However, Sujit Chakravorti, Federal Reserve Bank of Chicago, agreed with Schmidt’s point that this type of monetary incentive does not work for irrational actors, such as pedophiles, terrorists, and other perpetrators of payments fraud who are not motivated primarily by financial goals. Clearly, these types of actors present a problem to society that goes far beyond payments. Some argue that the existence of such issues with broad implications for our society leads to the need for public sector intervention in the problem of payments fraud.

The role of the public sector

Payments markets contain strong public-good components. Gene Amromin, Federal Reserve Bank of Chicago, argued that payments services are neither purely public goods nor purely private goods; thus, they are best provided by the private sector but with government oversight. Because of the inherent conflicts of interest, as noted in the previous discussion concerning misaligned incentives, the public sector can help counter information asymmetries by designing proper mechanisms to deter fraud, helping to align incentives to prevent fraud, and providing information to all levels of the payments system about the issue of payments fraud. While government involvement might therefore be seen as a crucial component in combating payments fraud, no clear consensus emerged at the conference on the best specific strategies for doing this.

Charles Docherty, MBNA Canada Bank, offered a perspective on how other nations deal with the role of government in payments fraud. In Canada, where there are fewer financial institutions and the central bank is not an active participant in the payments market, payments issues are largely governed by the private Canadian Payments Association, which consists of credit unions and banks. Docherty argued that in Canada, consumers and payments providers are considered the first line of defense for fighting payments fraud, followed by the government.

In contrast to the payments environment in Canada, in the United States regulatory and legal incentives have always been a central aspect of payments. Christian Johnson, University of Utah S. J. Quinney College of Law, noted that there are four types of laws that directly affect how payments fraud issues are handled (most of them involving the public sector): contracts between payments parties; state laws and regulations; federal laws and regulations; and international laws and treaties. All participants in the payments system must recognize these legal constraints.

Greene highlighted the importance of the government in the extremely crowded and competitive U.S. payments market. He said that the payments industry is concerned that sharing data and strategies related to payments fraud prevention might be viewed as collusive, possibly leading to a need for objective government intervention. Richey noted that by setting uniform rules, the public sector would be in a unique position to get at the root of payments fraud. However, Richey cautioned that too much intervention would stifle innovation. Some audience members argued that a uniform set of standards for all payment channels, governed by one body, would greatly deter payments fraud.

Ledig commented that the recent proposal by U.S. Treasury Secretary Henry M. Paulson, Jr., to give the Federal Reserve more power over all payment forms would be a step toward centralizing payments policy. Charles Evans, president and chief executive officer, Federal Reserve Bank of Chicago, reiterated that one of the key responsibilities of the Federal Reserve is to maintain the integrity of the U.S. payments system. Malphrus suggested that even in the current framework, which does not give the Federal Reserve governance over the entire payments system, the Fed should take up both advisory and participatory roles for that system. Such a role would still let the private market thrive. Some in the audience suggested that the Federal Reserve is in a unique position to advise on payments fraud issues, since it is both a direct participant and an overseer of the payments marketplace. Others, however, argued that these roles could prove conflicting for the Fed. Overall, conference participants seemed to favor a balanced approach of government and central bank intervention with support that would still allow the private market to police itself.

Conclusion

Participants in the conference felt that some level of fraud will always remain: Fraud could be eliminated entirely from the market only by shutting down active payment channels. However, a consensus was reached that the effects of data breaches and information compromises can be minimized through a holistic approach to data security. Such an approach would recognize the importance of cooperation within and across companies and among various actors in the private market. This cooperation would also be advanced by
government actions that are able to bring more uniformity to fraud mitigation without stifling innovation.

Fraud is an ongoing issue in the payments market, and the fast pace of technological change is likely to bring new opportunities for fraud to occur at the same time that it will spur more efficient fraud mitigation solutions. Policy leaders around the globe are struggling to define new rules and expectations of market participants, and industry leaders have different perspectives on the state of payments fraud and its future. The articles included in this volume represent various views on payments fraud from academic and industry speakers at the Federal Reserve Bank of Chicago’s 2008 Payments Conference.

NOTES

1Identity theft is another aspect of payments fraud. However, when payments information is used to help criminals obtain information about consumers in order to commit identity theft, the crime goes beyond payments. We do not focus on identity theft in this article.


3Conspirators obtained the credit card and debit card numbers by hacking into TJX Companies’ wireless computer networks. At the time, TJX Companies was in the process of becoming compliant with the Payment Card Industry Data Security Standard (PCI DSS), which defines guidelines for merchants’ handling and processing of payment card data in order to prevent card fraud and data breaches. See Brad Stone, 2008, “Global trail of an online crime ring,” New York Times, August 11, available at www.nytimes.com/2008/08/12/technology/12theft.html. Also see www.privacyrights.org.


5A phishing attack uses randomly distributed emails to attempt to trick recipients into disclosing personal information, such as account numbers, passwords, or Social Security numbers. See www.spamlaws.com/online-credit-card-fraud.html.

6In March 2007, the Federal Reserve Bank of Minneapolis held a roundtable discussion on payments fraud. A variety of market participants and regulators participated in the discussion. At this roundtable, participants revealed varying levels of comfort with the current state of payments fraud. See Board of Governors of the Federal Reserve System, 2007, “A summary of the roundtable discussion on retail payments fraud,” report, Washington, DC, July.

7Phishing is explained in note 5. During a phishing attack, a hacker tampers with the domain name resolution process so that users might go to the website of a legitimate financial institution and be unknowingly routed to a compromised site, where they reveal their personal information. A skimming device is one that is mounted to an automated teller machine or point-of-sale machine to copy encoded data from the magnetic stripe on the back of a payment card. For more information, see www.spamlaws.com/online-credit-card-fraud.html.

8Perimeter security refers to security systems that are developed to stop criminals from getting inside a network or database. In a SQL injection attack, a hacker uses knowledge of the SQL programming language to obtain hidden information in a database or network.


14For some details on the Check Clearing for the 21st Century Act, see www.federalreserve.gov/paymentsystems/truncation/.

15A neural network is a system of programs and data structures that mimics the neurons in the human brain. Neural networks “remember” information and data in complex ways. See www.webopedia.com/TERM/N/neural_network.html.


17For a more detailed argument for an increased governmental role in payments, see Stacey L. Schreft, 2007, “Risks of identity theft: Can the market protect the payment system?,” Economic Review, Federal Reserve Bank of Kansas City, Fourth Quarter, pp. 5–40.

18Ronald Mann, Columbia Law School, was originally slated to moderate the panel on fraud loss and dispute resolution. Christian Johnson moderated in his absence.

2008 Payments Conference
Payments Fraud: Perception Versus Reality

Thursday, June 5, 2008

INTRODUCTION AND WELCOME
Gordon Werkema, First Vice President and Chief Operating Officer,
Federal Reserve Bank of Chicago

KEYNOTE SPEECH
Divided We Fall: Fighting Payments Fraud Together
Mark Greene, Chief Executive Officer, Fair Isaac Corporation

IDENTIFYING SECURITY ISSUES IN THE RETAIL PAYMENTS SYSTEM
Moderator: Robert Ledig, Partner, Fried, Frank, Harris, Shriver & Jacobson LLP

Panelists
David Poe, Managing Director, Edgar, Dunn & Company
Ellen Richey, Chief Enterprise Risk Officer, Visa Inc.

Talking Points
What are the main security threats to retail payments?
What are the potential costs of payments fraud and of solutions to guard against it?
What role, if any, should public authorities play to protect payments system participants
from these threats?

FRAUD CONTAINMENT
Moderator: Bruce Summers, Payment System and Technology Management Consultant

Panelists
Jeff Schmidt, Consultant
Bob West, Chief Executive Officer, Echelon One
Mallory Duncan, Senior Vice President and General Counsel, National Retail Federation

Talking Points
What are the most common forms of retail payments fraud?
What are the most effective fraud reduction tools, and how have these tools evolved
to support “real-time” payments?
How do payment providers and merchants balance fraud risk and consumer convenience?

FRAUD LOSS AND DISPUTE RESOLUTION
Moderator: Christian Johnson, Professor, University of Utah S. J. Quinney College of Law

Panelists
Mark Michelon, Senior Director, E-commerce Risk and Revenue Protection, Orbitz Worldwide
Duncan Douglass, Partner, Alston & Bird LLP
Charles Docherty, Legal Counsel, MBNA Canada Bank
Talking Points

Who is responsible for mitigating fraud in the payments system, and what are the consequences of that responsibility?

How are losses allocated when fraud occurs?

Do current fraud resolution measures distort incentives for payments system participants to adequately secure payment information?

Friday, June 6, 2008

WELCOME AND INTRODUCTION

Daniel G. Sullivan, Senior Vice President and Director of Research, Federal Reserve Bank of Chicago

SECURITY RISKS AND SOLUTIONS IN EMERGING PAYMENT CHANNELS

Moderator: Bruce Cundiff, Director of Payments Research, Javelin Strategy and Research

Panelists

David Walker, President and Chief Executive Officer, Electronic Check Clearing House Organization (ECCHO)

Paul Tomasofsky, President, Two Sparrows Consulting LLC

Kevin Fu, Assistant Professor, University of Massachusetts Amherst

Talking Points

Do new payment channels, such as mobile, electronic images of checks, and decoupled debit, entail different fraud risks?

Are new tools necessary to minimize risks associated with emerging payment platforms?

Do these new channels provide any security features that mitigate risk in the payments system?

KEYNOTE SPEECH

Introduction: Charles L. Evans, President and Chief Executive Officer, Federal Reserve Bank of Chicago

Steve Malphrus, Staff Director for Management, Board of Governors of the Federal Reserve System

PUBLIC AND PRIVATE RESPONSES TO PAYMENTS FRAUD

Moderator: William Roberds, Research Economist and Policy Advisor, Federal Reserve Bank of Atlanta

Panelists

Judith Rinearson, Partner, Bryan Cave LLP

Allison Edwards, Director of Product Development, Fiserv EFT

Marsha McClellan, Chief, Money Laundering and Asset Forfeiture, United States Attorney’s Office for the Northern District of Illinois

Talking Points

How can the government define its role in fraud containment without stifling innovation?

Should different payment instruments have similar laws and regulations governing them?

Have standards been an effective tool in combating payments fraud?

CLOSING REMARKS

Sujit Chakravorti, Senior Economist, Federal Reserve Bank of Chicago
As consumers and merchants increasingly adopt electronic payments, the pricing of these services has generated substantial scrutiny around the world. Some public authorities have directly intervened in the payments market. Others have relied more heavily on the private market to develop payments pricing strategies. Moreover, innovative vehicles and business models may increase competition, resulting in greater choice to payments system participants. However, these products may not provide the same benefits as traditional debit and credit cards. In light of these developments, the Federal Reserve Bank of Chicago will host its ninth annual Payments Conference, Payments Pricing: Who Bears the Cost? During this two-day event held at the Chicago Fed on May 14–15, 2009, we will focus on the following:

- Evaluating the role of public intervention;
- Comparing perspectives on market-based solutions;
- Offering incentives to affect payments behavior;
- Leveraging technology to increase competition; and
- Developing future payments pricing strategies.

Who should attend?
Decision-makers from financial institutions, payment networks, regulatory bodies, central banks, merchants, and payment innovators, as well as academics.

Registration
Early Registration Discount—before March 6, 2009: $350
Registration from March 6, 2009, through May 1, 2009: $450
Registration received after May 1, 2009: $550
Cancellations after May 1, 2009, will not be refunded. Please notify us in writing of any substitutions.

Hotel Reservations
Union League Club of Chicago  
65 W. Jackson Blvd.  
Chicago, IL 60604  
(312) 427-7800  
www.ulcc.org

Hotel Blake  
500 S. Dearborn St.  
Chicago, IL 60605  
(312) 986-1234  
www.hotelblake.com

For more information, please visit  
www.chicagofed.org/paymentsystems  
or contact Susan Parren at (312) 322-4021 or susan.parren@chi.frb.org.
Fraud containment

Bruce J. Summers

Fraud is an unfortunate aspect of the technical efficiency of the payment system, which is measured by the quality of its operational performance and cost. Fraud degrades operational performance and increases cost—not only for the parties to the transaction(s) whose payments are disrupted, but for the payment system as a whole. Indeed, any serious consideration of payments fraud must account not only for the readily measurable business and consumer impacts of such fraud, but also for impacts on the performance and cost efficiency of the payment system.

Today's panel on fraud containment has been asked to identify the most common forms of retail payments fraud; the most effective fraud reduction tools, especially those pertaining to real-time payments; and approaches that payment providers and merchants take to balance fraud risk and consumer convenience. In taking up the last issue in particular we will attempt to provide a broad perspective that addresses the consequences of fraud not only for individual businesses and consumers, but for the integrity of the payment system as a whole.

While the focus of the conference is, naturally, on the U.S. payment system, it should be noted at the outset that the fraud problem is global, affecting many national payment systems and cross-border payment arrangements. For example, payment system fraud poses a threat to the internal market for payments in the European Union and is therefore receiving prominent attention in Europe. My sense is that the main payment system fraud concerns and issues in the U.S. and Europe are very similar and that we have a lot to learn from each other's experiences and responses. Accordingly, we should consider today's discussion part of a global dialogue about payment system fraud, and we should be open to opportunities to be informed by the international debate. This is especially so with regard to the public policy responses to the fraud problem.

The members of the panel bring an ideal combination of informed perspective and practical experience to bear on the problem of fraud. We have an information security technologist, a banking security practitioner, and a seasoned retail industry lawyer who has been concerned with customer data privacy and protection. Each of the panelists, whom I will introduce in a few minutes, will take 15 minutes to present his perspectives, and then we look forward to taking your questions and engaging in dialogue with you.

I would like to begin with some introductory remarks intended to set the stage for the panel discussion. In particular, I want to crystallize the business and public policy issues that involve containment of retail payments fraud. I will do so by summarizing the thinking of practitioners (by which I mean the providers and corporate users of payment services) and economists about fraud and efforts to contain it. The views of these two groups vary somewhat and are important because they influence public policy. You will understand that my background as a central bank economist, and also as a payments product manager and technology manager, has a strong bearing on how I cast the issues.

Economists' view of payments fraud

Payment system economists are principally interested in the most effective and efficient possible operation of the payment system. Of course, economists also respect the role of markets in delivering efficient
outcomes, and the payment system market is no exception. From the perspective of economic analysis, however, payment systems and markets are thought of as special because they entail something called “network effects” and “two-sided” services, which are characteristic of public goods.  

Payment markets, moreover, may not always function like perfect markets because of the presence of “externalities,” meaning that the costs and/or benefits associated with payment services are not always recognized by the parties to commercial transactions. As an example, my decision to use a risky means of payment may be a relatively easy one if it imposes costs on others and on the payment system, but not so much on myself. In addition, the markets may suffer from “asymmetrical information,” meaning that the sellers and buyers of payment services are not equally well informed about the riskiness of a particular payment service. For example, as a buyer I may not know as much as I would want or need to know about how well my personal payment information is secured in the service provider’s systems.  

For these reasons, as I will describe later, some economists see a natural role for the public authorities in helping control payment system fraud. They might do so by issuing regulations that specify the amount and type of disclosure required for payment service security, by enforcing those and other regulations, and possibly by facilitating industry-wide practices that lead to desired effectiveness and efficiency outcomes for the payment system.  

The views of economists are often informed by observed experience, and accordingly, I would like to share with you some lessons learned by practitioners who have met the business challenge of delivering effective, efficient, secure, and well-controlled payment services, especially as it pertains to security. They have found, first, that security is hard to achieve and ensure, and it requires relentless attention. Second, security is very expensive to produce and can impose indirect but nonetheless very real costs on consumers through the “user experience.” Third, cooperation across the supply chain is not only desirable but also necessary to achieve meaningful outcomes for customers because security is “only as strong as the weakest link,” as the adage goes. Fourth, certain aspects of technology, and security in particular, are moving outside the banks’ sphere of core competency, leading to outsourcing as a means of staying ahead of the curve; this leads to new types of risk that must be managed. Finally, the reputational risk associated with providing payment services is of greatest consequence to boards of directors of banking institutions because the success of the banking franchise depends on reputation and trust. Any business consideration of fraud containment must start with the board of directors and the corporate culture surrounding the private market approach to fraud containment.  

The Federal Reserve’s role  

A word or two about the Federal Reserve’s operational responsibilities in the payments marketplace will help illustrate that the Fed is in close touch with business and operational realities faced by practitioners. The Federal Reserve Banks directly provide retail payment services, primarily check, electronic check, and automated clearinghouse (ACH), for which they charge fees that are designed to recover the full costs of operation. They also produce retail payment services on behalf of the U.S. Department of the Treasury in their role as fiscal agents. This includes electronic payment services in support of the Treasury’s public debt and, if I can put it in these terms, accounts receivable and payable operations. The Fed thereby indirectly interacts with a large proportion of the retail public. Moreover, and perhaps especially important in the context of today’s discussion, the Federal Reserve Banks’ electronic payment operations are Internet-intensive, meaning that the public Internet figures prominently in the delivery of their services.  

This brings the reality of public networking and protection of customer information very close to home for the Federal Reserve Banks. Speaking of close to home, this is an opportunity to recognize the leading role played by the host of this conference, the Federal Reserve Bank of Chicago, as the Reserve Bank responsible for the content, quality, security, and bottom line financial viability of the Federal Reserve’s electronic payments. The Chicago Fed deserves to be recognized as the U.S. central bank’s Internet payments pioneer.  

An industry perspective of payments fraud  

The current state of thinking by industry practitioners about retail payment system fraud is well represented by the diverse cross section of participants in a 2007 roundtable on the subject sponsored by the Federal Reserve Board’s Payments System Policy Advisory Committee.  

The roundtable, which included representatives of banks, nonbank payment providers, card companies, and technologists, produced a variety of views but also a broad consensus on some important points. There was consensus that the current level of payments fraud is being effectively managed and that organizations must constantly adapt to keep pace with criminal activity, technology-driven change, and innovation in the payment system. At the same time, the industry representatives concluded that it
will never be possible to eradicate fraud completely and that the never-ending challenge of fraud prevention must balance costs and benefits.

While the roundtable participants indicated that the dollar value of fraud relative to business revenue is declining, their business costs of fraud mitigation are both substantial and trending upward. An especially interesting consensus emerged: The payment instrument that is the principal source of fraud losses on a comparative basis is the traditional paper check. We should try to validate this observation today and, depending on the outcome, reflect on the implications for future fraud containment as reliance on electronic payments continues to increase.

The roundtable participants spoke to the challenges posed by the Internet as a source of fraud, since it allows fraud that is directed to the domestic payment system to originate anywhere in the world. Some took a broad view of payments fraud by saying—rightly so in my view—that protecting customer information is part of the responsibility shouldered by payment providers. In the end, it was noted that detecting and preventing retail payments fraud requires a holistic approach that includes not only designing and producing well-secured payment services, but also encouraging and helping customers to practice good security behaviors. The roundtable made three suggestions for improving fraud detection and prevention. These are to increase 1) industry-wide information sharing and collaboration, 2) use of enhanced authentication technologies, and 3) adoption of the standards set by the PCI (Payment Card Industry) Security Standards Council LLC.6

The consensus reached by the roundtable is supported by the results of a somewhat earlier survey of approximately 100 large nonfinancial firms that actively use a variety of payment services.7 In the survey, each firm identified its most important payment processing needs and those needs that are least well met. While the firms participating in the survey generally responded that controlling fraud is very or critically important, a relatively low percentage responded that they are dissatisfied with the ability of current payment methods to control fraud. Consequently, other payment improvements, such as the ability to track transactions, emerged as needing higher priority attention than fraud containment.

Public versus private responses to payments fraud

The evidence suggests that practitioners are comfortable with the current state of fraud control in the retail payments marketplace. Their views can be contrasted with those of economists who take a public policy interest in the payment system. Economists’ current thinking about retail payment system fraud is somewhat more difficult to discern than that of practitioners because it has a work-in-process quality to it. Nonetheless, some recent economic analysis suggests that the view of economists is likely to be a bit less sanguine than that of the practitioners in the retail payment industry.

There seems to be the sense that market incentives and mechanisms per se are not up to the task of containing fraud and possibly other operational risks to a degree that optimizes overall payment system effectiveness and efficiency, and indeed they might not even maintain the integrity of the payment system as it continues to evolve. Two recent economic analyses undertaken within the central banking community suggest that the growing role of third-party, or nonbank, providers of payment services is a cause for concern and, moreover, that the public-good aspects of payment systems call for a more active governmental role. Let me elaborate briefly on some of the main conclusions from these analyses for they are important.

The role of nonbanks

A paper presented at the recent conference on nonbanks and risk in retail payments, sponsored jointly by the European Central Bank and the Bank of England, shows that nonbanks currently play an important role, especially in the United States, and will play an increasingly important role in a variety of retail payment systems worldwide.8 It argues that the growing nonbank presence has increased operational risk, including data security risk and, by extension, fraud risk. The paper also raises concerns about systemic operational disruptions as a consequence of concentrating operations among fewer key nonbank payment services providers. Finally, the paper speaks to the “payment system gatekeeper” role of banks and to the inherent difficulties that banks have in fulfilling their role while the operational locus shifts to nonbanks.

I think that it is very useful to measure and highlight the significant and increasing role of nonbanks in the retail payment system. At the same time, however, I question the conclusion that a more prominent operational role for nonbanks automatically increases operational risk. Electronic payments are among the most technology-intensive financial services. My practical experience with electronic payments is that the pace of change in the technology environment, including the technical capabilities that support fraud schemes, requires providers to operate on or near the technology frontier, especially if they want to stay a step ahead of the bad actors who perpetrate fraud.
Staying a step ahead of payments fraud in this environment is simply not possible for banks to accomplish without forming business alliances and partnerships that mobilize the needed technology skills. These business partnerships more often than not take the form of outsourcing to nonbank specialists, which, if managed well, act to strengthen the payment system. Also, I question whether concentrating the supply of sophisticated operational services, at least up to a point, necessarily increases operational risk. I think, again based on practical experience, that fragmented operations poorly performed, or performed below a recognized high standard, can be riskier than consolidated operations performed at the highest standard if due attention is given to security, business continuity, and operational contingency arrangements. Of course, operational cost is also a factor in that electronic processing exhibits natural economies of scale.

Information-dependent transactions

An additional paper relevant to the topic of fraud containment is that by a Federal Reserve Bank of Kansas City economist regarding the ability of the private sector alone to protect against the risk of identity theft and to protect the retail payment system. This paper focuses on “transactional identity” and “information-dependent transactions” involving noncash retail payments. It concludes that because of the problems with externalities and asymmetric information, the marketplace will not contain identity theft to an efficient degree; and as a result, the integrity and efficiency of the payment system, which we are to think of as a public good, are threatened. The concept of market failure is evoked and an active role for public authorities is envisioned to ensure the integrity of the payment system. Some examples of public policy prescriptions to deal with market failure—such as disclosure rules to address the asymmetric information problem and laws to clearly and comprehensively assign liability to address the problem with externalities—are very familiar to us.

The paper holds out the more intriguing prospect of other payment system interventions by public authorities along the lines of the Federal Reserve’s lender of last resort role in the credit markets or the federal deposit insurance.

This economic analysis seems to be at odds with the views of industry practitioners who think that the payments fraud challenge, while significant, is within the power of the private sector to address. The challenge, I think, is to evaluate seriously what remains to be done in the realm of private sector initiatives to protect the integrity of the payment system, not just the integrity of individual service offerings.

Conclusion

As we head into the panel discussion, it will be important to keep in mind the apparent differences in how practitioners and payment system economists size up the problem of fraud, the ways in which it is contained, and the implications for public policy. In taking up the issues assigned to us—the most common types of payments fraud, the most effective tools to deal with these types of fraud, and the costs of containing fraud—the panelists will provide their business perspectives and also help us understand whether the private sector is able to do enough alone to contain fraud in a manner that protects the payment system as a whole. The issue of the integrity of the payment system becomes more important each day, as electronic real-time payments supplant conventional paper instruments, dependence on sophisticated technologies increases, and nonbanks come to play an increasingly important role as providers of payment services. Depending on the outcome of the debate, public policy institutions such as the Federal Reserve could come to play a more active and interventionist role in the payment system as regulators and supervisors, and nonbanks could come more directly under the regulatory and supervisory purview of the authorities.
NOTES


2 This panel, which I moderated, comprised Jeff Schmidt, an independent consultant; Bob West, chief executive officer, Echelon One; and Mallory Duncan, senior vice president and general counsel, National Retail Federation.


6 Experience shows that continuous and timely strengthening of recommended standards deserves as much emphasis as does their adoption. For example, it has been reported recently that the standards set by the PCI Security Standards Council provide incomplete protection of “data in transit” through telecommunications channels. See Associated Press, 2008, “Credit card breach raises broad concern,” *New York Times*, March 23, and Joseph Pereira, 2008, “Credit card security falters,” *Wall Street Journal*, April 29.


Data security, privacy, and identity theft:
The economics behind the policy debates

William Roberds and Stacey L. Schreft

Introduction and summary

A byproduct of improved information technology has been a loss of privacy. Personal information that was once confined to dusty archives can now be readily obtained from proprietary data services, or it may be freely available (and, as Facebook users know, often voluntarily provided and accessible) through the Internet. While the increased collection and dissemination of personal data have undoubtedly provided economic benefits, they have also diminished people’s sense of privacy and, in some cases, given rise to new types of crime.

Is this loss of privacy good or bad? Press accounts repeatedly argue the latter: Too much data are being collected in ways that are too easy for criminals to access. But in a thought-provoking essay, Swire (2003) argues that a meaningful answer to this question requires some notion of efficient confidentiality of personal data—that is, of a degree of privacy that properly balances the costs and benefits of our newfound loss of anonymity. In this article, we explore the concept of efficient confidentiality, using some ideas from economic theory.

Loss of privacy: The costs are large and easy to find

The most dramatic consequence of the increased availability of personal information has been the emergence of a new form of payment fraud, identity theft. The 1998 U.S. Identity Theft and Assumption Deterrence Act (ITADA) defines identity theft as the knowing transfer, possession, or usage of any name or number that identifies another person, with the intent of committing or aiding or abetting a crime. Traditional varieties of identity theft, such as check forgery, have long flourished, but over the last decade, identity theft has become a major category of crime and a significant policy issue.

Identity theft takes many guises, but it is divided into two general categories: existing account fraud and new account fraud. Existing account fraud occurs when a thief uses an existing credit card or similar account information to illicitly obtain money or goods. New account fraud (traditionally) occurs when a thief makes use of another individual’s personal information to open one or more new accounts in the victim’s name. Both types of identity theft depend on easy access to other people’s data.

Today, identity theft is big business. A study conducted by the Federal Trade Commission (FTC), encompassing both new account fraud and existing account fraud, indicates that in 2006 identity thieves stole about $49.3 billion from U.S. consumers. When the time and out-of-pocket costs incurred to resolve the crime are added in, identity theft cost U.S. consumers $61 billion in 2006 (Schreft, 2007). Even this is a conservative estimate, however, as it omits certain categories of identity theft and some types of costs that are not generally known to consumers. For example, an increasingly prevalent type of identity theft is fictitious or synthetic identity fraud, in which a thief combines information taken from a variety of sources to open accounts in the name of a new fictitious identity (Cheney, 2005; and Coggeshall, 2007). There is no single victim, in contrast to traditional types of identity theft, but retailers and ultimately consumers end up bearing the cost.

Much of the data used in identity theft is obtained through low-tech channels. In consumer surveys,
victims who know how their identifying information was stolen commonly attribute identity theft to stolen wallets or mail or to personal acquaintance with the identity thief (Kim, 2008). In these same surveys, however, the large majority of identity theft victims are unable to pinpoint how the thief obtained their data. Available evidence suggests that much of these data are obtained through illicit access (called “breaches”) of commercial or government databases.

Statistics on data breaches are available from information security websites, such as Attrition.org and the Identity Theft Resource Center (www.idtheftcenter.org). Certainly data breaches are numerous and increasing: Attrition.org lists 326 reported data breach “incidents” for 2007, leading to the compromise of 162 million records of personal data, as compared with 11 reported incidents and 6 million compromised records in 2003. These numbers must be placed in perspective. A data breach does not necessarily lead to identity theft, and one reason for the upsurge in reported breaches is the spread of state laws that require consumer notification when a data breach occurs (Anderson, Durbin, and Salinger, 2008). Nevertheless, there is widespread recognition that data breaches promote identity theft. A strong demonstration of this can be found in the August 2008 indictment of an 11-person, global identity theft ring, responsible for the theft of 41 million credit card and debit card numbers, as well as hundreds of millions of dollars in fraud losses.

The benefits of loss of privacy:
More subtle, but substantial

If identity theft costs the U.S. economy so much, are there offsetting benefits? To try and make sense of this question, we will employ a branch of economics known as monetary theory. Broadly speaking, monetary theory seeks to understand how transactions are structured within an economy.

The classic model of monetary theory was proposed by Knut Wicksell (1935). Wicksell’s model economy is depicted in figure 1 and consists of only three individuals: Andy, Bob, and Clyde (A, B, and C, for short). Andy can produce a good valued by Bob, Bob can produce a good valued by Clyde, and Clyde can produce a good valued by Andy. The point of Wicksell’s model is that in real-world economies, transactions typically happen between people who cannot deal through simple barter. For example, when Andy and Bob meet, Bob would like to purchase Andy’s good, but the good that Bob has available to trade is only valued by Clyde. How should exchange proceed?

One way to solve this problem is through the use of cash. Suppose that A and B meet on every Monday, B and C meet on every Wednesday, and A and C on Fridays. Then if everyone agrees that the goods they exchange are each worth $1, the economy can function perfectly well with a “money supply” of two dollar bills. For example, Bob sells his good to Clyde on every Wednesday, earning a dollar that he uses to buy Andy’s good the following Monday. Andy uses this dollar to buy a good from Clyde every Friday, and so on. This “Wicksell triangle” shows how cash can function as a sort of recordkeeping system for transactions within an economy; every dollar that someone spends is proof of an earlier sale by the same person.

Cash has some well-known limitations, however—some of which appear even in the context of this simple model. For example, if Clyde gets sick or otherwise fails to show up one Wednesday, then Bob will have no money with which to make next Monday’s purchase. In practice, cash has other drawbacks, including risk of counterfeit or theft, the inconvenience of finding an automated teller machine (ATM), limited usefulness in telephone and Internet transactions, and the fact that it does not pay interest.

The alternative to cash is credit. In Wicksell’s model economy, cash would not be needed if A, B, and C could get together and agree that each individual would receive a good of their preferred type, so long as they had provided a good to someone else the previous week. That way, if an individual occasionally was unable to sell his good during one week, he could still purchase goods on the expectation that he would resume sales the following week. The value of this additional exchange of goods, beyond what would be possible if all transactions were only in cash, is known as a credit benefit.

FIGURE 1

Wicksell triangle

Note: See the text for further details.
Paying by credit has other advantages, which are the “mirror image” of the disadvantages of cash: fewer trips to the bank, less liability in case of theft, ease in transactions at a distance, and the reduced need to carry non-interest-bearing cash.\footnote{8}

Any estimate of the total credit benefit in an economy is somewhat speculative, since it involves the comparison of the value of exchange in an actual economy to the value of exchange in a hypothetical economy where only cash is available for most transactions.\footnote{9} For a developed economy such as that of the U.S., however, this benefit is almost certainly quite large. For example, in 2006 (the year of the FTC identity theft survey), U.S. residents made about $3 trillion in purchases, using credit and debit cards.\footnote{10} If the credit benefit of these transactions alone (ignoring other types of credit transactions) amounted to, say, just 5 percent of their total value, the resulting benefit to the overall economy would be $150 billion—more than enough to outweigh the estimated costs of identity theft.

In the rest of this article, we will argue that some loss of privacy is central to the provision of this credit benefit.

**Identity: Real and transactional**

In Wicksell’s model economy, there’s no chance of identity theft. Andy, Bob, and Clyde are well known to one another, and as long as one of their mutual friends (say, Dave) can keep a tally of who’s provided a good to whom, it would be easy to maintain a credit-based system of exchange. This system would be self-enforcing, since any shirking by one party would quickly be noted by Dave and immediately become apparent to the other parties.\footnote{11} Such informal credit systems are common among friends and families, in primitive societies, and in other settings with limited social interactions.

But most transactions in today’s economy are either between 1) parties who are total strangers, and/or 2) parties who feel no particular sense of obligation toward one another. Credit in such situations requires some system to control two types of risk. The first type of risk is credit risk—the risk that the purchaser may not repay the debt incurred. Overcoming credit risk requires a way to keep track of “credit histories,” that is, a way to restrict the use of credit to people who habitually pay their bills. The second type of risk is fraud risk—the risk of deception by the purchaser. Overcoming fraud risk requires a way to associate transactors with credit histories: For example, I may have a spotless credit record, but somehow that information has to be conveyed to the grocery store before I’m allowed to leave the store with a bag of groceries. To be effective, both types of services require the accumulation, storage, and distribution of large amounts of personal data. But the data required by the second service concern a person’s identity, and are bound to be of a more confidential and controversial nature.\footnote{12}

“Identity” in general refers to all the distinguishing attributes of an individual—potentially a very long list. The term personal identifying data (PID) is used to describe some portion of a person’s identity—name, birth date, Social Security number, etc.—that is readily observable by others. In order to distinguish individuals, the credit bureaus, credit card companies, data brokers, and other parties in the credit industry have compiled large databases of PID. These subsets of a person’s “real” identity that are stored by these parties and used in transacting can be thought of as transactional identities (Schreft, 2007). Once the relevant data have been verified, a person’s transactional identity may be augmented by the creation of new, synthetic data unique to that person, such as a credit card number, PIN (personal identification number), and so on (Kahn and Roberds, 2008).

A typical credit transaction—say, a purchase of a bag of groceries, using a credit or debit card—can be thought of as a merchant exchanging goods in return for two essential pieces of payment information corresponding to the types of risk described previously: 1) that the purchaser, based on his credit history, is likely to pay his bill\footnote{13} and 2) that the purchaser’s transactional identity is genuine so that the consumer is not a fraudster.

**Transactional identities as club goods**

All credit-based payments require systems for processing valuable information. We can think of this information (credit histories and transactional identities) as economic goods, or items having value in exchange. These goods have value, since they facilitate the exchange of other goods (say, groceries) that people want to consume. Electronic versions of payment data, once amassed, can be stored at a few locations and then shared among payment system participants at very low cost. The data used in credit-based transactions meet Varian’s (1998) description of a digital good, a good that can be stored and transferred in digital form.

Digital goods are also nonrival goods, meaning that they are not diminished by successive use. This distinguishes them from rival goods, such as cars and cornflakes; one individual’s consumption of a rival good diminishes or eliminates the possibility of another person consuming it. Other examples of digital (and also nonrival) goods are given by the electronic information that is incorporated into broadcast and...
cable television, computer programming, or recorded music and video: For instance, my consumption of an episode of American Idol does not diminish another’s enjoyment of the same episode. The same holds true with payment data, including transactional identities: The fact that Wal-Mart knows that I am not a fraudster does not diminish the value of the same information to Home Depot.

Nonrival goods are classified as club goods or public goods. A club good is an excludable nonrival good—that is, one for which a group or individual can be excluded from consuming (for example, cable TV programming). A public good is a nonexcludable nonrival good—that is, one for which access cannot be limited (for example, national defense or clean air). The club good classification is more appropriate for payment information, since access to this information can be controlled (to a greater or lesser degree).

The “nonrivalness” of electronic payment information is a tremendous source of economic efficiency. Turning the clock back several decades, in any retail situation involving credit, a merchant had to independently come by the information needed to assess a customer’s creditworthiness. The high cost of this information meant that credit was impractical in many situations, for example, during travel or for small transactions. The development of the credit industry (the large databases of credit histories and transactional identities, credit and debit cards, electronic authorization procedures, and antifraud technologies) has meant that merchants can take advantage of economies of scale in managing this information, and has spread the costs of information management over a larger group of merchants (and, ultimately, consumers). This has, in turn, increased the credit benefit available to society as a whole.

Of course, the transformation of payment information into a nonrival good has not occurred in isolation. All kinds of data (music, video, maps, encyclopedias, and celebrity gossip) have been widely digitized, and thanks to the essentially nonrival nature of digital goods, they are rapidly accumulated and widely disseminated.

The dark side of nonrivalness

A central feature of any digital good is its quality. Recorded music or video, for example, is useless if the original is garbled. A potentially interesting website may seem less so if it is known to harbor computer viruses. Quality is especially critical for payment data because people using a payment system expect it to work flawlessly virtually 100 percent of the time. Contamination of a payment system’s data through even a few errors or instances of fraud can quickly erode its value.

A “dark side” of the efficient production of payment information is that it can compromise quality; that is, it can facilitate fraudulent activity as well as legitimate use. Once a fraudster has assumed another person’s transactional identity (through either new or existing account fraud), the fraudster becomes an apparently legitimate participant in one or more payment systems and, by extension, a legitimate participant in the eyes of many participants in those systems. This vulnerability means that payment data, as an economic good, will only have value in the presence of the complementary good “data integrity,” which is the quality and reliability of the data incorporated into the payment system (Braun et al., 2008). Data integrity, like the underlying payment data, is a nonrival (club) good: The assurance that a payment information database is secure against data breaches is not diminished by successive use.

Another widely recognized drawback of modern payment arrangements stems from the more difficult to measure, but nonetheless important, consequences of diminished privacy. That is, the digitization of personal data contained in transactional identities has made these data available to many more people than ever before, often with negative consequences. These may take the form of intangible, but undeniable, costs in terms of people’s loss of a “sense of space” about their personal lives. Or, for victims of identity theft, these costs may assume a more concrete form, through harassment by bill collectors, misplaced civil lawsuits, or even criminal investigations.

Many current payments and credit practices can be interpreted as attempts to partly restore the sense of privacy that may have existed in earlier times. When someone makes a purchase with a credit card, for example, that purchaser must effectively reveal some information to the merchant concerning his transactional identity—at least in the form of a relatively anonymous credit card number. This “surrender” of information represents a compromise between the merchant’s need to identify the purchaser and the purchaser’s desire to preserve his own privacy. Ideally, the merchant obtains enough information about the purchaser to determine that the transaction is legitimate, but no more. Consumers themselves have also undertaken forceful actions to safeguard their privacy, removing their names from public directories and mailboxes, installing paper shredders in their homes, and only giving out personal information to the most trusted parties.

Ironically, these very attempts to restore privacy may have contributed to the rise of identity theft, according to LoPucki (2003). LoPucki points out that in earlier times, individuals’ access to credit often depended
on their public persona, that is, on their standing within a local community or circle of business associates. Those seeking access to credit had to sacrifice much of their privacy (say, by socializing with their neighbors on a regular basis or joining civic organizations) in order to gain a reputation as an upstanding and creditworthy individual. Modern information technology, by enabling “instant credit” between relatively anonymous parties, has reduced the need for a public persona, but it has also multiplied the potential for fraud.

**Efficient confidentiality: Beyond supply and demand**

Using the ideas outlined thus far, we can now look at the issue of efficient confidentiality. The term confidentiality has a specific meaning in our context, which is the likelihood that a person’s transactional identity will not be observed by miscreants and put to inappropriate use. A person’s confidentiality can be thought of an economic good, whose provision in the marketplace depends on two other economic goods: 1) the amount of PID incorporated into that person’s transactional identity and 2) the level of security for these data, or the degree of data integrity applied to the person’s transactional identity. An increase in the second good always improves confidentiality. An increase in the first good can improve confidentiality, up to a point. The more data that are collected (all else being equal), the more precise the identification of individuals is, and hence, the greater the availability of credit-based payment is throughout the economy. But increasing the amount of PID collected (again, all else being equal) reduces privacy and can also amplify the negative consequences that occur when such data are misused, eroding confidentiality.

How should we know if these two goods (data collection and data security) are being efficiently provided? Textbook economic theory says that for many goods, it is (conceptually, at least) easy to describe how that good can be efficiently provided: An efficient market exists for a good when its supply curve intersects with its demand curve. The demand curve for a good, in turn, is given by its marginal benefit to buyers, and the supply curve is determined by sellers’ marginal cost of producing that good. In a competitive industry, if the price of a good is above (below) its marginal cost, producers enter (leave) the industry until efficiency prevails.

Unfortunately this familiar model doesn’t work for digital goods, since their marginal cost is practically zero. Instead, a more typical pattern for digital goods is for there to be competition among a few large producers, which are able to take advantage of the extensive economies of scale in these goods’ production (think of the computer software and entertainment industries). Prices remain above marginal costs, so as to defray the costs of production.

We see the same pattern in the construction of transactional identities by a relatively small number of large players such as credit bureaus, credit card networks, and card issuing banks. Through the accumulation of large amounts of PID, these organizations attempt to meet the demand for transactional identities that exists in the market economy. Just as with other digital goods, such as computer software and recorded video, it is hard to know whether these data are being efficiently collected and priced.

The situation is different when we turn to the issue of data integrity. Because payment data are only useful if they are communicated (in some form), these data must be touched by a large number of hands to be of any value. A real-world list of such hands would include consumers, merchants, credit bureaus, banks, and payment processors. In other words, efficient production of data integrity, a club good, requires the cooperative efforts of a large number of “club members.”

Large clubs often promote efficiency because they allow for economies of scale in the production of a good. But within large clubs, conflicts of interest can arise as to the amount of the good that should be provided. This is especially true for goods such as data integrity, for which the “weakest link” or “flood control” model of a nonrival good is often applicable.

For a weakest link good, the total amount of the good provided to the club is equal to the lowest amount of the good supplied by a club member (the weakest link in a chain, or the lowest levee in a flood control system). The idea of a weakest link is consistent with many press accounts of identity theft, in which a data breach at a single retailer or payment processor leads to widespread fraud. There is a natural tendency to supply an inefficiently small amount of a weakest link good (Varian, 2004), which can arise from the following conflict: A club member with relatively little at stake will tend to put less effort into providing the club good than a club member with a lot at stake. This tension is present in many situations involving data security (Anderson and Moore, 2006).

Recent changes in the payments industry’s security practices can be seen as a response to this problem. For example, a set of industry-wide data security standards—the PCI (Payment Card Industry) standards (www.pcisecuritystandards.org)—has been created as a way of strengthening the weakest links in the data security chain. Another development along these
lines has been the increasingly common practice of merchants quickly disposing of payment data, rather than storing it for an extended period of time. 19

An additional source of inefficiency comes from externalities (also called spillovers) across data security practices. An externality occurs when the consumption or production of a good by one party affects another’s, conferring benefits or costs on the other party. A negative externality results when a party does not take into account the full cost of his action to others.

In the context of data security, the potential for negative externalities exists for at least two reasons. First, as noted previously, payment data often passes through many hands, so it is difficult to determine how an identity thief was able to access the necessary data. Second, under current U.S. and Canadian laws, recovering the costs of a data breach through the courts can be difficult (Schreft, 2007; and Chandler, 2008). Either way, if payment data are stolen from one party and used to commit identity theft with costly consequences for another, the first party may not expect to pay the full costs of the breach. Taken together, these complications suggest that there are obstacles to the efficient provision of data integrity in the marketplace. Because payment system participants may not fully take into account all of the costs associated with their security practices, this can lead to underprovision of data security. This would, in turn, imply an inefficiently low level of confidentiality in the marketplace, even if the market is collecting the “right” amount of PID.

In Roberds and Schreft (2008), we present a model that shows how this inefficiency could be exacerbated by the interaction between PID collection and data security. If some payment systems are not adequately securing their data and other payment systems are alerted to this, then each system’s best safeguard against identity theft may be to increase the amount of PID it uses for transactional identities. Under these circumstances, gathering more PID can reduce fraud, but doing this is inefficient because it further reduces confidentiality.

**Roles for regulation**

The previous discussion points to a role for public policy. If the markets for information on transactional identities are providing inefficiently low levels of confidentiality, there may be ways for well-designed policies to improve on market outcomes.

One policy implication that is not supported is government entry into the markets for payment information. As with other types of club goods, the excludability of payment information provides a profit incentive to motivate ongoing improvements in efficiency. But the production of club goods is rarely a straightforward business, and it is usually subject to extensive policy interventions. Electronic entertainment products, computer software, and various types of Internet content, to name just three examples, are frequent subjects of public controversy, legislation, regulation, and litigation.

This same general pattern is found in the markets for payment information. Various pieces of legislation and regulatory efforts have sought to address the “weakest link” and “spillover” problems identified before, but have stopped short of trying to micromanage industry practices. For example, the Fair and Accurate Credit Transactions Act of 2003 (commonly known as the FACT Act) seeks to increase the industry standards for minimally acceptable security practices. The FACT Act requires banks and other creditors to develop procedures to respond to account activity that could reasonably be interpreted as evidence of identity theft (“red flags”), but does not specify the details of how this should be done. 20

In the same vein, a number of state laws now require that consumers be notified whenever their data are breached. One motivation for this requirement is to enable quicker detection of identity theft by consumers. An equally important purpose for this requirement, though, may be to motivate better security practices by increasing the costs of a data breach (in terms of both dollars and reputation). A number of states have taken another tack, which is to allow consumers to limit or “freeze” access to their credit reports, that is, to limit access to information on their transactional identities.

A concern with this type of regulation is the cost of compliance. Since securing data is costly, perfect confidentiality of personal data cannot be an efficient outcome, and should not be a goal of sensible regulation. As outlined in this article, some amount of identity theft is inevitable given modern information technology. Eliminating identity theft entirely would not be possible without eliminating the efficient sharing of information at the heart of our modern credit and payment systems.

**Public goods**

Government intervention is traditionally viewed as beneficial when it yields public goods. One such good is “public security,” as is provided by the criminal justice system. The ITADA and various state laws have sought to discourage identity theft by imposing severe criminal penalties—a form of deterrence not available to the private sector.

The nature of identity theft puts limits on the effectiveness of criminal sanctions, however. By stealing
someone else’s payment data, an identity thief gains that person’s access to credit in largely anonymous situations, such as in purchases over the Internet. This same anonymity that benefits legitimate purchasers (in terms of access to credit with increased confidentiality) makes criminal prosecution of identity theft impossible in many cases—as when the identity thief is located in a different country from that of the victim.

Another noteworthy public good in this context is that of overall “confidence” in credit and payment systems. As discussed previously, people do not like to use payment systems without something close to 100 percent reliability. If incidences of identity theft and data breaches were to become sufficiently common, the result could be a loss of this public good—that is, a loss of confidence not only in the directly affected parties, but in credit-based payment more generally (Braun et al., 2008). One rationale for recent regulatory actions in the payments area is that, apart from the effects of any specific provisions, these laws and regulations demonstrate governments’ commitment to maintain a reasonable standard for confidentiality of payment information.

Conclusion

In this article, we have looked at the issue of confidentiality of personal information from the standpoint of economic theory. Some loss of privacy is necessary for the credit benefit, which is a key advantage of modern payment systems. By consolidating personal information into transactional identities, information technology now allows people to enjoy this credit benefit in circumstances that would have been unthinkable a generation ago.

The sharing of information on transactional identities is vital to the operation of these payment systems. However, this information sharing can facilitate fraud in the form of identity theft. Information sharing can also create conflicts of interest that may not be easily resolved through the operation of the marketplace. Thoughtful public policy should be aimed at resolving these conflicts and providing public goods. The ultimate goal of regulation should not be absolute privacy of consumers or complete suppression of identity theft, but instead the promotion of efficient confidentiality of personal information.

NOTES

1See, for example, Stone (2007), Swartz and Acohido (2007), Caruso (2007), and Dow Jones and Company Inc. (2008b).

2There are no time-series data on identity theft rates, but one measure of the extent of the problem is the how often the term “identity theft” shows up in press reports. Anderson, Durbin, and Salinger (2008) report 30 mentions of “identity theft” in U.S. newspapers in 1995; 2,000 in 2000; and 12,000 in 2005.

3This estimate is from a survey of consumers reported in Synovate (2007); for extensive discussions of this survey, see Schreft (2007) and Anderson, Durbin, and Salinger (2008).

4Of course not all data breaches are publicized, so these numbers are probably underestimated.

5See, for example, Stone (2008). For other recent data breach incidents, see Braun et al. (2008).

6If we increase “money velocity” by changing the order of transactions (say A and C meet on Wednesdays and B and C on Fridays), then a money supply of one dollar bill will be sufficient.

7Beginning with Kiyotaki and Wright (1989), this role for cash has been extensively developed in “search” models of money; Wright (2008) surveys this literature.

8For a detailed comparison of the costs of cash versus other forms of payments in certain retail settings, see Garcia-Swartz, Hahn, and Layne-Farrar (2006).

9Even for the simple Wicksell model, calculation of a credit benefit can be a challenging exercise. Taub (1994) shows that for this model, people can sometimes do just as well by keeping hoards of cash. However, Kocherlakota (1998) shows that in general an economy’s credit benefit will be a positive number.

10Bank for International Settlements, Committee on Payment and Settlement Systems of the Group of Ten Countries (2008). Use of a debit card can result in a credit benefit if the card is attached to a bank account with an overdraft privilege or line of credit.

11In some simple economies like Wicksell’s, Araujo (2004) shows that Dave may not be needed; mutual confidence that others will honor their obligations is enough to sustain credit-based exchange. Kahn and Roberds (2009) discuss how Wicksell’s model can be used to analyze various types of payment systems.

12Credit risk and fraud risk are often difficult to separate. For example, if a person applies for a credit card, runs up a bill, and then never makes a payment, then it may be hard to tell whether the person meant to commit fraud or just wasn’t able to pay. Or someone may refuse to pay for his credit card purchase, claiming the transaction was fraudulent; this practice is sometimes known as “friendly fraud.” Nonetheless, it is useful to conceptually distinguish between these two types of risk.

13There is an element of credit even with many transactions that are thought of as “instantaneous” (for example, debit card or Internet banking payments), since these do not settle instantaneously. In many card transactions, the card issuer assumes the “credit risk” that the card payment will not be repaid by the cardholder.
An economy of scale occurs when an increase in the production of a good lowers its average cost. In our context, the increased accumulation and distribution of payment information have lowered the average cost of accessing such information.

A complementary good is defined as a good that is consumed with a second good, for which an increase in the demand for the first good results in an increase in demand for the second. For example, cars and gasoline are complementary goods.

The structure of this industry has been changed by the emergence of data brokers (legal and illegal) and other entities that compile and trade PID obtained from other sources (Schreft, 2007).

REFERENCES


Perspectives on retail payments fraud

Steve Malphrus

Let me begin by saying that I am not here to lecture, but rather to learn. Today, I would like to talk about a couple of things. First, I would like to start with some themes that emerged from a roundtable discussion that the Federal Reserve held last year with industry leaders on emerging issues involving fraud in the retail payments system. This is important to the Federal Reserve. The outputs from the roundtable are used to direct the Federal Reserve’s research and inform its work. Thus, hearing your perspectives on those themes today is important. The second thing I would like to talk about is an area in which I have been doing research. These are the emerging trends in new account fraud detection for applicants on the Internet, where businesses are not physically present to authenticate the identity of customers. As everybody here knows, this is an area of growing interest throughout the banking industry.

Findings from the roundtable discussion on retail payments fraud

Let me start with the roundtable that the Federal Reserve sponsored last year. Fourteen industry experts—including merchants and representatives from payments system providers, financial institutions, and law enforcement organizations—participated. Overall, these leaders agreed that, although the current level of payments fraud is being effectively managed and does not represent a crisis, organizations must constantly adapt to keep pace with criminal activity and with changes in technology and payment methods. While the dollar amount of fraud relative to business revenues in the United States is likely declining, the costs associated with fraud mitigation are substantial and increasing. The roundtable discussions focused on four main themes: 1) the changing landscape of retail payments fraud, 2) current trends, 3) emerging concerns, and 4) areas for improvement in fraud detection and prevention. The following paragraphs sum up our discussions involving these four themes.

The changing landscape of retail payments fraud

Despite declining use of checks across the country, industry leaders find that the largest number of fraud attempts remains in check payments. Fraud losses are also highest for checks on a comparative basis with other payment methods. A number of participants stated that business losses resulting from check fraud are significantly higher than losses from noncheck payment types because checks are relatively easy to alter or forge, using readily available printers, scanners, and computer software.

Moreover, changes in the payments system and in criminal behavior have introduced additional risk. One key change in the payments system has been the proliferation of commerce conducted over the Internet. The Internet has created new means for criminals to gain access to consumers’ personal and financial information, and has facilitated the formation of extensive illegal networks through which criminals buy and sell this information without the limits of geography. Indeed, substantial Internet fraud operations are now linked to sites located in certain developing countries. The Internet has also accelerated worldwide information-sharing among criminals regarding successful fraudulent schemes, so that new fraud techniques now move quickly around the world. In addition, the growth in online commerce has led to an increase in the number of transactions in which merchants are not physically present to authenticate the identities of purchasers.

Steve Malphrus is the staff director for management at the Board of Governors of the Federal Reserve System.
That said, some changes in the payments system have helped reduce risk, such as faster clearing of check payments associated with Check 21 and check-to-automated-clearinghouse (ACH) conversion. Being able to clear payments more quickly can mean that a fraudulent check may be returned before a collecting bank makes funds available to the depositor. At a minimum, faster returns help inform banks and their customers that fraud is taking place. But some feel that ACH e-check payments may be more vulnerable to fraud than other ACH standard code categories, such as ACH transactions initiated via telephone. Concerns were also raised over the greater use of check images in the rapidly growing Check 21 environment, which could reduce the usefulness of some current check security features that may not survive the imaging process.

Further, criminals’ ability to adapt to changes in the industry’s practices in fraud detection and fraud prevention is a continuing challenge, as these lawbreakers continue to seek the path of least resistance. For example, as large merchants and banks develop new tools to detect and prevent fraud, criminals turn to smaller and medium-sized enterprises because they are less likely to have the resources to invest in fraud detection and prevention. Because fraud affects the entire financial industry, some feel that it is the duty of larger businesses and banks to reach out to educate and aid smaller organizations. Others suggest that we should raise the bar by increasing criminal penalties for fraud and prosecuting fraud more rigorously.

Current trends

It is becoming increasingly important for firms to protect consumer information. Industry leaders are concerned about the potential damage to their brands’ reputations in the event of a data breach. The industry has taken steps to protect consumers from fraud that may result from compromised information. Often, for evidence of fraud, banks and card networks monitor customer accounts that may have been compromised and then reissue cards when necessary. Some industry leaders argued that, although the storage of data is a potentially vulnerable point in the payments system, the extent to which compromised information has actually been used is relatively low.

In many instances, if consumer information is compromised and subsequently used to commit payments fraud, the consumer is not liable for the associated losses. Thus, while it is important to protect consumer data, it is equally important to develop tools to prevent the fraudulent use of data or to otherwise render data unusable. One example is phishing. While phishing is a current threat to the security of consumer information, many believe that the level of actual loss incurred from phishing has been relatively low in the aggregate. In some cases, education has been reasonably effective in preventing consumers from divulging information online.

In addition, it is important to differentiate between “payments fraud” and “identity theft.” While both are a crime, the ramifications of each are substantially different. The Federal Trade Commission (FTC) has defined the term “identity theft” as fraud perpetrated by 1) obtaining access to and illegally using a consumer’s existing financial information, such as a person’s credit card number or bank account number, or 2) illicitly obtaining identity information about a consumer to open new financial accounts in the consumer’s name. The roundtable participants generally agreed that the second part of the FTC’s definition should be considered “identity theft” and that the first part should be considered “payments fraud.” Some stated that the FTC report used an overly broad definition of identity theft, which has led to an overestimate of the true frequency of this type of fraud. Nevertheless, the consequences of true identity theft can be very significant for consumers. While actual financial losses might be low, the impact on a consumer’s credit record—and the time and effort required to correct that record—can be substantial.

Emerging concerns

As noted, criminals are continually searching for weaknesses in fraud detection and fraud prevention practices. Several participants said that the potential movement of check-based fraud to the ACH network is an area of growing concern for the industry. A fraudulent payment initiated with a check can move into the ACH system through a point-of-purchase (POP), back-office-conversion (BOC), or accounts-receivable-conversion (ARC) transaction. Since ACH has traditionally been used for recurring payments from trusted sources only, banks may not yet have robust tools in place to detect fraudulent ACH payments from other sources. Fraudulent checks that may be detected using existing tools might, therefore, go undetected if processed on the ACH network. This possibility is a particular concern to businesses that use check fraud prevention services, such as positive pay, that are not available for ACH payments. While a concern, fraud of this nature is, at present, relatively low.

The industry has only recently begun monitoring the movement of fraud across payment channels. Perhaps further study is required to understand how fraud is moving between paper and electronic instrument
or between different electronic instruments. Banks and businesses are looking to adopt a holistic approach to detecting and preventing retail payments fraud across the spectrum of payments systems. One participant described this approach as managing fraud at the “relationship” level—that is, at the level of an individual or a corporate client for a bank, and a customer for a merchant—rather than at the “product,” or payment instrument, level.

Moreover, the industry is concerned that the introduction of new payments instruments, such as prepaid cards, could increase fraud in the payments system. One participant noted that some of these cards can be easily reloaded with funds and can be used anonymously, making them effective vehicles for money laundering. Another stated that open-loop, reloadable prepaid cards could be a primary vehicle for fraud in the future, and others concurred that prepaid cards are a growing area of concern. We also discussed the security of mobile and contactless card transactions. On the one hand, payments made using these devices could be more exposed depending on their security features. On the other hand, the development of security enhancements, such as “dynamic” authorization techniques, for some payment devices can offer significant benefits. The hesitation in trusting emerging payments instruments may stem from the fact that their risks are not yet understood. Successful payments systems have historically had to put innovative systems into production and undergo a learning phase before the development of a fully mature risk-mitigation strategy.

**Areas for improvement in fraud detection and prevention**

At the roundtable, the most discussed suggestions for improving the industry’s ability to detect and prevent retail payments fraud were 1) increasing industry collaboration and information sharing, 2) using enhanced authentication techniques, and 3) adopting Payment Card Industry (PCI) standards.

Merchants and financial institutions could benefit from increased collaboration and information sharing across industries and within their own business sectors, including through the development of best practices in fraud detection. Firms need to not only detect fraudulent transactions in process but also prevent fraud’s initial occurrence by improving authentication at the point of sale. At the roundtable, the effectiveness of PIN (personal identification number) and chip technology was debated. Some stated that fraud rates on PIN debit cards are significantly lower than those for other payment types; as a result, they advocated the application of PIN security to card payments in general.

Chip technology has been widely adopted in other countries, and could prove to be a safer alternative to magnetic stripe technology for card-based transactions.

Roundtable participants also discussed the role of the Payment Card Industry program, developed jointly by Visa and MasterCard. Full compliance with security standards could help the industry safeguard consumers’ personal and financial information. The PCI program in particular could be helpful, but there are challenges for some organizations to become compliant with the PCI program. Nevertheless, compliance with the PCI program might be a good first step in securing consumer information, though other opportunities exist. For example, existing data privacy regimes generally apply to banks or merchants, while they exclude others, such as third-party service providers. These third parties have access to consumers’ personal and financial information. In order to improve the security of consumer information, it is desirable to expand data protection regimes with respect to both the types of payments and the types of organizations that are included.

Ultimately, the roundtable discussions returned to the refrain that criminals will continue to search for the fastest and easiest ways to commit payments fraud. Consequently, strategies for fraud detection and fraud prevention should be considered holistically, so as not to merely shift fraud from one payments channel to another. Industry leaders maintain that it is not financially feasible to prevent all payments fraud. Rather, businesses must make prudent, risk-based decisions that will yield appropriate returns relative to the investment required to minimize fraud. Organizations continue to balance costs and benefits when investing in tools to mitigate fraud.

At the roundtable’s conclusion, several suggestions emerged for how the Federal Reserve might assist the industry’s efforts to mitigate fraud. Some advised the Federal Reserve to continue its outreach events to encourage industry participants to share concerns and effective practices, and others emphasized the importance of the Federal Reserve conducting research on payments and fraud-related issues. As a general matter, however, leaders advocated the continued application of market-driven approaches to keep payments fraud at a manageable level. Payments system participants’ ability to adapt to changes in criminal behavior will be critical in maintaining a safe and efficient payments system.

**Some thoughts on new account fraud**

Shifting gears now, I would like to offer my perspective on recent developments in the detection of
fraud in new accounts. Many companies with an online presence today are struggling to find solutions for screening out fraudulent applicants for new accounts. These accounts range from those used for banking and brokerage accounts to accounts used for services. The dilemma is universal for online businesses where there is no person-to-person discussion with the applicant and, therefore, no possibility to examine documents such as driver’s licenses or passports and to verify identity in person. New account fraud in such non-person-to-person (mainly online) environments is estimated by some experts to be four to five times higher than it is when accounts are opened in person.

Although there is no comprehensive solution available in the market today, various methods can help detect accounts opened for illicit purposes. In the case of regulated banks, meaningful attempts must be made to detect new account fraud under the new “red flags regulations” that were fully implemented by November 2008 (I discuss these regulations in greater detail later). This is true whether the ultimate victim is a consumer, whose identity has been stolen, or the business itself, where an account is opened using a fictitious identity created by a criminal.

**Client device identification**

In the non-person-to-person online environment, a business does not have an opportunity to screen identity documents, videotape the person, and/or engage directly with the applicant. However, the business does have an opportunity to screen the user’s device, such as a personal computer. Various technologies make device identification and analysis a useful first step in flagging suspect applicants. For first-time users, businesses can obviously not rely on installed desktop software, tokens, or credentials that have already been installed. However, they can analyze various pieces of information available through the user’s web browser connection to check for potentially fraudulent activity. These include the following.

- **Geolocation** of the user based on the user’s Internet protocol (IP) address. Vendors that specialize in IP address intelligence are often able to detect the use of blacklisted IP addresses or blocks of addresses (that is, those that have been known to be used for criminal activity). They can detect the use of anonymizers and/or proxy servers that criminals use to hide their locations. Businesses can also compare the country and geographical region of the IP address to the country and region from the user’s credit card billing address.

- **Personal computer/web browser identification** examines the hypertext transfer protocol (HTTP) browser header and other information from the user’s computer or device, and compares them to what are expected. For example, this process can compare the time stamp from the computer to the time expected from the user’s geolocation. Using a JavaScript executed from the business’s server, this software can try to uniquely identify a computer and determine if it is being used by a large number of account applications. Software is available today that specializes in computer identification using proprietary techniques along with geolocation analysis. Similarly, a biometric system that records a user’s keystrokes and unique typing pattern can be used to ascertain if the same person, and not just the same machine, is opening multiple accounts.

- **Botnet detection** can identify a machine on a criminally run botnet that is accessing an enterprise’s website.

**Fraud detection using information on the account application**

Device identification tests can be subject to further fraud screening through the use of information entered on the application. Depending on the information requested on the application form, these fraud detection strategies can include the following.

- **Identity proofing**, which is typically used when a comprehensive set of information is being requested from the user, such as financial data, Social Security number, employment history, and homeownership information. This is common when applications are filled out for financial accounts, such as insurance, credit card, and bank accounts. Identity proofing can be relatively expensive, at a few dollars for each identity checked, and uses either:
  - Rule-based data-matching systems from vendors or credit bureaus; or
  - Identity scoring, relying on service and software providers that detect potential fraud using scoring models that look across application records and data.

- **Credit card fraud detection**, which is useful for new account openings that require only a credit card authorization. This detection typically costs about 15 cents to 25 cents per transaction, on top of the usual authorization costs, and depends upon
volume and vendor-licensing arrangements. These systems analyze data available from credit card records, such as billing address and shipping address. They perform various checks, such as validating addresses using the card companies’ address verification system, and compare credit card billing and shipping addresses to the customer’s geolocation and to lists of suspect addresses. The systems check to see how many times the end-user accesses a webpage asking for credit card information—possibly an indication of a brute force attack against a card’s security code. Credit card fraud detection systems also can compare credit card numbers provided by the user with stolen cards noted on blacklists, although stolen credit cards are so readily available to the fraudsters that blacklists have limited value. Most systems for credit card fraud detection enable enterprises to manage the business rules that each of their transactions runs against, so the businesses can catch fraud patterns particular to their situations.

- **Niche data verification**, which refers to the verification of specific data, such as telephone numbers or applicants’ ages. These data are then reconciled with data expected from the applicant. The line information database is a telecommunications industry standard database containing the same information made available through hub providers. Unfortunately, it is still not possible for enterprises to get access to a comprehensive set of wireless phone directories held privately by some wireless carriers (notably Verizon Wireless)—a step critical in verifying phone numbers because many customers prefer listing cellphone numbers rather than landline numbers.

### Stepped-up applicant verification

Optimally, all account applications should go through a set of initial screening procedures, and suspect transactions that need further review should be routed to a fraud investigation queue for manual or automated follow-up. Additional automated screening can occur using one of the following methods.

- **Identity proofing** is a method that uses knowledge-based authentication systems, based on public source data that pose questions to the user that only he or she can presumably answer (such as “What was the make of the first car you owned?”). Vendors offer identity-proofing applications based on public records, which can be partially effective in screening out fraud. However, roughly 20 percent of the question/answer sessions invoked for high-risk applications fail or are abandoned. Sometimes, the failure is because legitimate users cannot successfully answer the questions or because there is not enough public data available for a particular individual. At other times, criminals manage to answer questions successfully.

- **Telephone-based user verification** is a method that relies on a call to an applicant using a phone number found in the public records or provided by the user personally. The automated phone system can simply ask the user to speak, and it can record the user’s voice or ask the user to type in the phone transaction number generated by the online application session. This method is not foolproof unless the business is sure that the phone number on record belongs to a legitimate user.

### Implications of red flags regulations

On October 1, 2007, the Federal Trade Commission and federal banking regulators, including the Federal Reserve, the Office of the Comptroller of the Currency, the Federal Deposit Insurance Corporation, the Office of Thrift Supervision, and the National Credit Union Administration, released rules that require financial institutions to step up efforts to combat identity-theft-related fraud. These long-awaited identity theft rules implement the Fair and Accurate Credit Transactions Act, or FACT Act, and took effect on January 1, 2008. Financial institutions covered by the rules had until November 2008 to comply.

The rules require regulated financial institutions to create “reasonable policies and procedures” for detecting and preventing identity theft. Red flags cited in connection with an account application or an existing account include patterns of activity that are inconsistent with the historical and usual pattern of an account, such as a recent and significant increase in inquiry volume or an unusual number of recently established credit relationships. Other red flags include applicant addresses that do not match addresses from external sources, as well as internally inconsistent personal information, such as a lack of correlation between the Social Security number range and the date of birth. Institutions are also asked to check for invalid phone numbers or addresses and to flag applications for which an address, Social Security number, or phone number provided is the same as that submitted by other persons opening an account or by other customers.
Conclusion

As we ponder retail payments fraud going forward, the risk is not just about the cost of dealing with fraud and the associated losses. Indeed, fraud risks and associated retail payments fraud will cross into areas of public policy related to privacy. Today privacy is becoming a serious issue, and interestingly, this issue brings us full circle to the broader topic of information security.

Government agencies, for example, have a new mandate in terms of handling information about citizens: It is called private identity information. Federal agencies must take affirmative action to protect private information such as Social Security numbers, dates of birth, etc. Moreover, today the U.S. Department of Homeland Security has an assistant secretary for cybersecurity and communications. That position centralizes the federal government’s work in this area as well. Other agencies that work on privacy and identity issues related to payments fraud include the Central Intelligence Agency, the National Security Agency, U.S. Department of the Treasury, and the Federal Reserve System. Concerns about terrorist financing and money laundering drive much of this federal work, but we should remember that such concerns are also increasingly spilling over into the world of payments fraud. In the future, you should see additional coordination and partnerships between the public sector and the private sector to address risk.

I think it is important to understand that the Federal Reserve System is unique in that it acts as a banker’s bank, the federal government’s bank, and a payments system operator. Having a payments system that is safe and secure is an absolute necessity in maintaining the confidence and trust held in it. To achieve this, we must focus on operations risk first, but also pay attention to reputational risk. It is important for us to understand these risks from multiple perspectives—from the economic research perspective, from the perspective of a financial market authority, and from the perspective of a very large bank.

NOTES

1 For details on the Check Clearing for the 21st Century Act, see www.federalreserve.gov/paymentsystems/truncation/.

2 A phishing attack uses randomly distributed emails to attempt to trick recipients into disclosing personal information, such as account numbers, passwords, or Social Security numbers. See www.spamlaws.com/online-credit-card-fraud.html.

3 Positive pay is an antifraud service offered by virtually every U.S. commercial bank. It protects a company from altered checks and counterfeit check fraud by comparing the components (such as the account number, check number, and dollar amount) of each check presented for payment against those from a list of checks previously authorized and issued by the company. It allows a company to reject unauthorized transactions (that is, for checks that do not match) before losses occur.
It is a great pleasure to be addressing this august group. As some of you know, I began my career at the Federal Reserve back in 1982. So speaking to you is like a homecoming for me. I have been fortunate in my career to participate in the U.S. banking economy from three perspectives: at the Fed, obviously a policymaking central bank; at Citibank, a lender; and at two financial technology providers, including 12 years at IBM (International Business Machines) and the last year at Fair Isaac, a leader in decision management technology.

From these three perspectives, I have seen the tremendous collaboration that exists in the banking industry on the issue of fraud. However, from my current vantage point, I am also able to see a disturbing trend: More companies are declining to participate in some of these collaborative, consortium-based best practices. The reason is simple: They see a competitive advantage to keeping their information and experience to themselves. This raises some key issues for the financial services industry.

Do we want to fight fraud or move it around? That is, do we want to reduce the amount of fraudulent activity overall, or are we content to just have the most advanced banks move it to the less advanced banks, and to shift it from well-protected channels to less protected channels? Does a failure to maximize our effectiveness at fraud prevention have even deeper consequences? Which people, which groups, and which activities might we be funding if we allow fraud to persist? And are private industry initiatives enough, or is there a role in fraud prevention for public sector initiatives, mandates, or intervention?

I won’t leave you guessing as to where I’m going with this. My experience has taught me the following.

- Individual gains must be balanced by the collective good.
- It is better to stop a fraudster than send him to the bank next door.

Now, my company is in the business of giving banks a competitive advantage. We have used consortium approaches to defeat fraud. We believe these collaborative approaches, along with ubiquity in protection, are essential ingredients in the fraud-fighting formula. They are necessary to reduce the “balloon effect” in fraud prevention, where progress in fighting a segment of fraud succeeds primarily in moving fraud from one place to another. We win when fraud loses—and fraud loses when we fight it together.

**Types of payments fraud**

Let me start by simply defining the key areas of payments fraud I’m discussing here. Fundamentally, we can divide fraud into two categories. There is first-party fraud, which is the abuse of account privileges by the account holders themselves, or the acquisition or expansion of those privileges by deceitful means. There is also third-party fraud, which is often identity fraud, or the abuse of one person’s account by another. For the purposes of this talk, I am not discussing insider fraud, which is the misuse of a customer account by bank employees or others involved in the provision and distribution of financial services products.

First-party fraud typically involves your customer opening an account with you, with the intention of violating the terms of the account agreement. It can also involve a borrower selling his information to...
criminals or constructing a fraudulent identity or deceitful credentials for gaining credit. This type of fraud very often shows up in the collections queue as bad debt. But it is not traditional bad debt—when it is intentional, it is fraud.

Third-party fraud is what we usually think of when we consider fraud. This is stolen identities, the use of lost or stolen cards, and the counterfeiting of cards or other means of account access. It encompasses a wide range of techniques. This is where the criminal gangs operate—and where advanced technology comes into play to greatly reduce fraud losses.

**Fraud costs**

Fraud—both first-party and third-party—is on the rise, but not across the board, according to Javelin Strategy and Research. That is because fraudsters are fast learners and attack less protected channels. Almost 4 percent of adult Americans were victims of fraud in 2007, resulting in losses of $1 billion. U.S. credit card fraud losses were down 22 percent to $11.4 billion; credit cards are highly protected by consortium models that are part of the Falcon fraud protection system. (I will talk more about that later.) By contrast, U.S. debit card losses rose 16 percent to $7 billion. Debit card transaction volumes are on the rise, and only some debit cards are protected by consortium Falcon models. Online purchase fraud experienced an increase, rising 33 percent in 2007. Though new account fraud incidents increased, total annual new account fraud losses dropped by 21 percent. There was a surge in new telephone account misuse, and existing checking and savings accounts fraud was up by 10 percent.

Just to take one example of a rising problem, card-not-present fraud (CNP fraud) is on the rise (see figure 1). It is estimated that about half of transactional card fraud today is CNP fraud. CNP fraud is primarily perpetrated through fraudulent use of cards for online purchases. CNP fraud is the biggest threat to online channels, such as PayPal.

Looking at global card fraud, we can see how the different methods of fraud have been changing over time. Certain fraud types are rising to “fill the gap” made by excellent progress in categories such as lost or stolen card fraud, since new technologies and channels enable new forms of abuse, as demonstrated by the rise in CNP fraud. To summarize, I have noted the following:

- We can make a huge difference by focusing on fraud in a collaborative way; and
- Fraudsters are moving from one channel and technology to the next, in what we call the balloon effect—squeeze them in one area and they move to another.

So are we winning the war on fraud, or just moving it around? We don’t need any help recognizing the importance of fraud in its impacts to our businesses and the bottom line. But it is worth noting that real economic costs may be 150 percent of measured fraud losses. In other words, we are underestimating the problem when we just measure fraud losses. We know from our work with clients, for example, that a tremendous amount of bad debt is actually misclassified fraud. We worked with one prominent UK card issuer and found that more than 10 percent of the bad debt in its collections queue was really fraudulent activity. The costs associated with this is not just the charge-off losses; it is also the costs of having collections and recoveries staff and agencies try to collect unrecoverable monies.

**Fraud’s shifting focus**

Of course, the costs to the lending institution and its customers are not the only costs we need to worry about. Terrorists and criminal organizations are funding crime through fraud. The costs here are incalculable. These costs make a strong case that a concerted, collaborative effort to fighting fraud is more important.
than making fraud prevention a competitive advantage for a select group of lenders.

I’ve mentioned the balloon effect in fraud. The fraud detection and prevention tools that have been commonly applied by banks include card issuer and network transaction fraud solutions, a debit bureau and other identity protection for account opening, the implementation of chip and PIN (personal identification number) technology, the increasing usage of account verification techniques, and online fraud detection and transaction review tools. However, many new types of fraud have emerged or increased in response to the banks’ defenses. These include the following:

- Increasing phishing and skimming attacks;
- More attacks on small card issuers and smaller merchants that do not have the same level of protection;
- Recruitment of insiders to better enable fraud;
- Offshore fraud;
- Mail theft of cards;
- Large-scale abuse of card data retained at the point of sale;
- Declining effectiveness of address verification in detecting fraud; and
- International mail-order, telephone order, and online fraud.

The point here is that gains in one area of fraud are frequently offset by losses in another.

In fact, banks, retailers, telecommunications firms, and others are struggling to combat fraud, which is growing more complex all the time. There are more channels and lines of business to protect. There are regulatory mandates for better risk management. We are fighting sophisticated, worldwide criminal organizations. There are more frequent pattern changes. Lost, stolen, and counterfeit cards remain a concern, but we are also dealing with new forms of attack, such as Internet attack bots, which apply all kinds of techniques—persistence being the key ingredient—to work their way through online security measures.

**Fraud solutions**

Mass compromise losses could rocket higher given the low current criminal utilization rate of compromised cards. Large data breaches, to date, have been inefficiently leveraged by the criminals that end up with the information. Some incidents involving thousands of card numbers have resulted in only a few handfuls of fraudulent transactions. But breaches perpetrated by a more organized or effective criminal organization could have much more severe and immediate consequences.

The uneven protection of account types has raised interest in enterprise fraud systems. The information in figure 2 comes from a survey of leading U.S. banks conducted by TowerGroup for Fair Isaac this year. These banks are pursuing enterprise fraud systems as a way of controlling fraud losses. Today’s fraud systems tend to protect one channel or product. It is like putting a burglar alarm on your front door but leaving the windows open. An enterprise fraud system is like a burglar alarm system for your whole house. This sounds simple, but it isn’t. Few institutions today have the same level of protection across the organization. There are a lot of very well-protected doors out there—and some very open windows as well. As we discuss the importance of collaboration, it is important to understand that many of the principal victories that have been made in the area of fraud depend on collaboration. Next, I present three examples and focus on the collaborative aspect.

**Falcon Fraud Manager**

How does collaboration win today? It probably comes as no surprise that I’m starting with Falcon
Fraud Manager, which is a Fair Isaac solution. Falcon is an excellent example of the effectiveness of collaboration in fighting fraud. Falcon is the leading cards fraud protection platform. Falcon manages 65 percent of card accounts worldwide, including 90 percent of credit cards in the U.S. Falcon reviews card transactions and “scores” them based on their likelihood of being fraudulent, enabling card issuers to stop losses faster and to react dynamically to changing fraud activity in real time. Falcon’s fraud detection is based on innovative neural network models that are “trained” on large sets of consortium data. These consortium models are embedded in end-user software or accessed by card issuers via third-party processors. The neural network models search through masses of data to identify very subtle signs of fraud. The size and diversity of the data are critical factors in the power of the models. We have created a fraud consortium that includes information on 1.8 billion card accounts, contributed by lenders that subscribe to the Falcon product.

Falcon Fraud Manager typically cuts individual issuers’ fraud losses by 50 percent and in many cases by more. But the really impressive thing is the impact this kind of solution can have on the industry. Falcon Fraud Manager was introduced in 1992, when card fraud was at 18 basis points in the U.S. As shown in figure 3, this number has since declined by about two-thirds based on the industry’s use of a common, powerful fraud protection system.

This shows how a ubiquitous solution powered by close collaboration has served to benefit both individual issuers and the industry. Individual issuers have squeezed fraud out of their portfolios, and the industry as a whole has worked to squeeze a substantial amount of fraud out of the system.

Card Alert

Our second example involves automated teller machine (ATM) fraud detection. Some 11,000 banks in the U.S. subscribe to a Fair Isaac service known as Card Alert. What Card Alert does is trace the flight path of compromised cards to identify compromised ATMs. It works backward from compromised cards to identify whether they passed through a single ATM. The Card Alert team then identifies other cards that passed through the ATM in question. They notify the issuers that these cards may be at risk. The system currently flags roughly 500,000 unique card accounts annually as being compromised at ATM devices. So the information on card compromise from some issuers is used to benefit other issuers and to help criminal investigators.

Card Alert generates a wealth of data on ATM fraud trends, which is used by banks to systematically stop the fraud and by law enforcement to fight the fraudsters. Collaborative efforts like Card Alert have served to dramatically reduce the percentage of fraud that occurs at ATM devices.

Chip and PIN fraud

Our third example looks outside the U.S., to the chip and PIN rollout in the UK. This was an industry-wide, collaborative effort that resulted in nearly all devices being PIN-verified in the UK and, therefore, nearly all cards being much harder to counterfeit or scam. Over 90 percent of UK cards are chip and PIN cards now, and nearly 1 million retail tills have been upgraded. In 2005, this resulted in a 24 percent reduction in fraud from counterfeit, lost, and stolen cards, according to APACS (Association for Payment Clearing Services) in the UK.

However, while counterfeit, lost, and stolen card fraud has been contained by chip and PIN technology, it has pushed fraud for those same accounts to a new venue. Cross-border fraud—largely unprotected by the chip and PIN technology—went up by 43 percent in 2006 and by another 77 percent in 2007. Cross-border

![Figure 3: Card fraud in the U.S., 1990–2006](image-url)
fraud now accounts for 39 percent of all fraud for UK card issuers, compared with 27 percent in 2006. This shift swallows nearly all of the gains achieved through the reduction in fraud occurring in the UK itself. The problem that we’re seeing here is that the collaboration worked in the UK, but because it was not executed in easily accessible neighboring countries, it failed to reduce UK issuers’ overall losses. They decreased one form of fraud but increased another. Again, this speaks to the importance of both collaboration and ubiquity in avoiding the balloon effect.

**Device and merchant profiling**

How will new technical advances enable the industry to combat fraud? Let’s look at three new advances that use payments data in different ways to increase fraud protection. The first is known as device profiling. One of the ways that successful card fraud solutions operate is to build a profile of each cardholder that can be used to identify unusual activity. By profiling devices as well, we are able to provide a more complete profile picture for a given transaction. The device profile looks for unusual device behavior: large amounts, rapid transactions, and suspicious patterns of transaction types.

Device scores can be combined with the cardholder scores to improve fraud detection. This approach can identify patterns that often involve multiple cards. It is especially useful in identifying counterfeiters and ATM burst fraud events. Device profiling requires a collaborative cross-issuer view, similar to the Card Alert service discussed before.

Our research shows a sizable predictive lift from adding cross-issuer device profiling. For example, there is an 80 percent relative performance lift in real-time value detection at a 10:1 false positive rate. This means that at a threshold where you are flagging ten “good” accounts to review for every one fraudulent account, you are identifying 80 percent more fraud than a traditional card system based on just cardholder profiles. This kind of trade-off curve looks geeky to you, you have to understand that I am an econometrician working at a company populated by analytic staff. Geeky is where I work!

Our second innovation involves merchant profiling. As we discussed, today the standard is to profile cardholders and use every transaction to build and evolve the profiles. What we can do now is build a fuller picture by examining the merchant profiles as well. Merchant profiles are similar to cardholder profiles in that they contain a summarized view of detailed transaction information and history. They identify the points of sale that are more or less likely to experience fraud.

The account fraud score is adjusted downward or upward based on the merchant information. This additional data collection increases the detection power of the model, through the integration of cardholder variables, merchant variables, and combined cardholder/merchant data. Better fraud detection means lower losses and improved customer service. Again, the ability to profile merchants effectively depends on the rich data coming from a cross section of issuers.

We have found that using merchant profiles in Falcon Fraud Manager, our card fraud system, enables clients to jump another level up in fraud detection. The enhanced version of Falcon, that is, with merchant profiles added, identifies substantially more frauds in real time, enabling the issuers to reduce fraud losses. At that same 10:1 false positive rate, the consortium subscribers are able to achieve a 40 percent relative performance lift in fraud detection and prevention.

**Adaptive models**

Our third example is a different kind of technology breakthrough. It involves what we term “adaptive models.” The fraud models we have been discussing so far are based on consortium data, and every year we update the models by training them on the most recent set of consortium data. These new models are then used to upgrade our clients’ systems. This has been very successful, but it means there is a lag time between the card issuers’ experience of evolving fraud trends and the incorporation of that experience into their fraud-fighting tools. What we need is a way to capture new and important shifts in fraud patterns because of the highly dynamic nature of fraud.

The way adaptive models work is to adjust the model weights on each issuer’s system. This dynamically tunes the models in response to actual fraud experienced by the issuer. This approach enables the issuer to benefit both from the broader view of fraud activity captured in the consortium model and from more immediate information on fraud against their accounts.

Our ability to detect fraud is increased with the adaptive models. Our research has shown an 18 percent relative performance lift in real-time value detection at a 10:1 false positive rate. These are just some of the advances coming in payments card protection. The point is that to make these kinds of advances, and to make them effective, requires collaboration. I have pled my case regarding collaboration.

**Collaboration**

What are the implications for the industry? The real frontline soldiers in the war on fraud—in particular the fraud managers who help protect their
institutions from a growing array of threats—need the best weapons we can give them. The innovations they depend on often stem from independent action and proprietary development. But these innovations are powered by collaboration. The trend toward viewing fraud management as a competitive advantage has potential negative implications for fraud management overall.

Models are stronger when they are trained on larger, more varied data sets. Certain types of information, such as device profiles, only provide value when powered by a macro-level view. And because fraud always finds its way to the weakest link in the chain, ubiquity helps contain the problem of the balloon effect.

So where might the public and private sectors collaborate next? Here is one idea: an industry-wide Fraud Alert Network. This would take the success of systems such as Falcon and Card Alert to a new level by building on collaboration. A Fraud Alert Network could take an approach to updating systems that is similar to the way companies such as AVG and Symantec fight computer viruses. By looking across millions of events, they are able to identify new virus patterns and automatically push updates to their user bases.

This is the model we are exploring for payments fraud. Rather than annual system or model updates, we would push out updates, rules, or hot lists automatically.

The concept includes a collaborative rules subscription service, as well as simplified, timely consortium data collection. And the Fraud Alert Network includes a portal designed to bring banks, retailers, and others together to share ideas. Think of it as a private user community focused on real-time fraud issues—a Facebook for fraud management. In fact, this collaboration portal will go live later this month. We expect it to yield faster responses to fraud threats. It is a great example of where we see fraud protection going—toward greater collaboration and a real unified front. In summary, I leave you with these key ideas.

- Payments fraud remains a front burner issue.
- Fraud evolves with new payment product technologies.
- This is too big an issue to fight separately.
- Private sector collaboration is essential, as we have seen—it is really the foundation of the successful antifraud initiatives.
- Public sector involvement can help with best practices and information sharing.

In short, this is a war—divided we fall, united we win.

NOTES

1 A phishing attack uses randomly distributed emails to attempt to trick recipients into disclosing personal information, such as account numbers, passwords, or Social Security numbers. A skimming device is one that is mounted to an automated teller machine or point-of-sale machine to copy encoded data from the magnetic stripe on the back of a payment card. For more information, see www.spamlaws.com/online-credit-card-fraud.html.
An examination of the fraud liability shift in consumer card-based payment systems

Duncan B. Douglass

Introduction and summary
In the absence of a significant (and right now unforeseeable) shift in the retail payments landscape in the United States, consumers will continue to reach consistently (and often) for their debit and credit cards. They will use these cards when paying for goods and services in face-to-face, Internet, mail order, and telephone order transactions. Likewise, criminals will continue to use tried-and-true tactics and will develop innovative methods to perpetrate payment card fraud.

At the intersection of consumers conducting legitimate card transactions and fraudsters pursuing their illegal ends is a tangled web of public laws and private card network rules. These laws and rules allocate fraud risk among the consumers, card issuers, and merchants participating in card-based payment systems. In theory, one would hope that these laws and rules for payment card transactions are thoughtfully designed to encourage behavior that minimizes fraud losses to the system as a whole. In reality, systemwide fraud reduction is often not the principal objective behind particular public laws or private rules affecting fraud liability allocation. Consequently, these laws and rules may fail to promote efficient fraud avoidance; indeed, in some instances, they may actually discourage fraud avoidance.

Defining the issue
The first step in evaluating the efficiency of fraud liability allocation rules in current card-based payment systems is to define the issue. Doing so requires an understanding of the difference between identity theft and common payment card fraud, as well as an understanding of the workings of the card-based payment systems at issue.

Identity theft versus fraud
News stories abound about identity theft resulting from dumpster divers absconding with old bank statements and criminals rifling through mail and intercepting credit card offers. Further, email accounts are barraged with phishing attempts and other web-based schemes craftily designed to lure consumers into revealing personal identification information that can be used for nefarious purposes. Typically, the fraudsters intend to use the ill-gotten fruits of their snooping to impersonate their victims and access their credit or asset accounts. This is identity theft, and it is an increasingly pervasive problem in the United States and throughout the world. During 2007, Consumer Sentinel, a network that collects information about consumer fraud and identity theft from the Federal Trade Commission and over 125 other organizations, recorded 258,427 identity theft complaints.

Identity theft is distinguishable from common financial fraud. Identity theft is generally defined as “the use of personal identifying information to commit some form of fraud.” In contrast, fraud is simply “[a] knowing misrepresentation of the truth ... to induce another to act to his or her detriment.” As noted in the definition of identity theft, fraud is typically the end goal of identity theft. However, often fraud is committed without antecedent theft of Social Security numbers or other assumption of identity. Along with the cases of identity theft reported in 2007, 555,472 cases of non-identity-theft-related fraud were reported during the same year. Given that card-based payment systems (and other payment systems, for that matter) seek to prevent monetary fraud perpetrated through the system regardless of how the information used to perpetrate the fraud was obtained, here I focus on the broader category of payments fraud—whether or not...
it is precipitated by identity theft. There is no need to steal another person’s identity to perpetrate simple payment card fraud—all the perpetrator needs to do is obtain a person’s payment card or payment card information.³

Distinguishing fraud from identity theft is important to the discussion that follows for two reasons. First, fraud is broader and more pervasive than identity theft. Second, the means of preventing fraud in the initiation of payments, and the appropriate allocation of losses that result from payments fraud, are generally not dependent on whether the fraud resulted from identity theft or from a simpler card/data theft incident. There is no doubt that consumers who fall victim to identity theft experience significant nonmonetary losses in addition to the losses resulting from the fraudulent transactions. These include the opportunity costs of time spent disputing fraudulent claims, closing existing accounts, and opening new accounts.⁶ However, public laws and private rules governing card payment systems are not capable of preventing such costs to consumers because these costs are wholly external to the payment system itself.

**Payment systems fraud generally versus signature-based card fraud**

Having distinguished identity theft from payments fraud and clarified that this discussion is concerned with the latter, it is worth making the distinction between payment systems fraud generally and payment systems fraud perpetrated through means of a signature-based access device. This distinction is important because public law treats access device fraud differently than other types of payment systems fraud. Moreover, private card network rules related to fraud are generally different for signature-based card products than for other payment products (including card products based on a PIN, or personal identification number). For the purposes of this article, I limit my consideration to signature-based consumer debit cards (which are directly or indirectly linked to, and draw funds for settlement from, a consumer asset account) and credit cards (which are linked to, and draw funds for settlement from, a line of credit extended by the card issuer). These types of debit and credit cards are issued for acceptance on the major credit card networks in the United States: Visa, MasterCard, American Express, and Discover.

Of course, there are other payment card forms and other types of accounts that can be accessed using payment cards. These include wireless technology key fobs, biometric account access that uses no card at all, and prepaid cards that access a different type of account altogether. Again, I only discuss signature-based debit and credit cards here because these devices and the accounts they access remain the most prevalent in the retail payment systems marketplace.

**Allocation of payment card fraud liability: Public laws and private rules**

Determining which party to a given fraudulent payment card transaction has liability for the fraud requires an understanding of both the applicable public legal framework and the private card network rules. A fundamental assumption in this article (and many others, although the point is often unstated) is that the actual wrongdoer—the perpetrator of the fraud—will be unavailable for recovery, and so one of the innocent parties involved in the transaction must be asked to bear the resulting loss. Absent any public laws or private rules to the contrary, the cardholder would be the risk bearer by default unless a benevolent merchant or card issuer agreed to absorb the loss. Luckily enough for cardholders, both public laws and private card network rules intervene to protect cardholders and to reallocate liability for fraud losses among other participants to a fraudulent card payment transaction.

**Public law**

The public law framework that serves to protect consumer users of credit and debit cards from bearing the full brunt of fraud losses associated with lost or stolen access devices are as follows: the Truth in Lending Act (TILA), together with Regulation Z, and the Electronic Fund Transfer Act (EFTA), together with Regulation E.⁷ Historically, Congress has shown a fair degree of restraint in tinkering with TILA and the EFTA. Instead, Congress has allowed the Board of Governors of the Federal Reserve System to use its regulatory authority to extend appropriate consumer protections to new payment products and account structures through revisions to Regulation Z and Regulation E.⁸

Likewise, the Federal Reserve Board generally has taken a measured approach in amending Regulation Z and Regulation E to address market developments (for example, transactions initiated by mobile phone) and new funding sources accessed by payment cards (for example, prepaid accounts held by the card issuer in an omnibus account structure).⁹ The Federal Reserve Board expressly acknowledged its restrained approach to expanding regulations when it promulgated the interim final rule extending Regulation E coverage to payroll cards, noting that the Board was not extending coverage more broadly to prepaid cards because “coverage of such products could impede the development of other card products generally.”¹⁰
Truth in Lending Act and Regulation Z

Under TILA and Regulation Z, cardholder liability is capped at $50 for all unauthorized transactions, regardless of whether the fraud occurs in a single transaction or multiple transactions and regardless of when the cardholder learns of the loss or theft of the card or reports the loss or theft to the card issuer. Regulation Z defines unauthorized use in connection with a credit card as use “by a person, other than the cardholder, who does not have actual, implied, or apparent authority for such use, and from which the cardholder receives no benefit.” Unauthorized use of a credit card includes both physical use of a lost or stolen card or fraudulent use of information from a credit card, whether or not the actual device has been lost or stolen. Thus, fraudulent use of a credit card number and expiration date to conduct a card-not-present transaction over the Internet constitutes “unauthorized use” according to Regulation Z.

Electronic Fund Transfer Act and Regulation E

The EFTA and Regulation E place a floating cap on a consumer cardholder’s liability for unauthorized debit card use under which the maximum liability amount is determined when the cardholder notifies the card issuer of the loss or theft of the card used to perpetrate the fraud. If the cardholder notifies the card issuer within two business days of learning of the loss or theft of the debit card, the cardholder’s maximum liability is limited to the lesser of the actual amount of unauthorized transfers or $50. If the cardholder fails to notify the card issuer within two business days of learning of the loss or theft, the cardholder’s maximum liability is capped at $50 for all unauthorized transactions.

In addition, if the cardholder fails to notify the card issuer of unauthorized activity within 60 days after the card issuer sends a periodic statement reflecting the unauthorized transactions, subject to the $50 and $500 liability caps, the cardholder has unlimited liability for fraudulent transactions occurring after the 60th day.

It is worth noting that negligence of the cardholder in safeguarding the debit card is not a basis for the card issuer to impose greater liability on the cardholder than is otherwise permissible under the EFTA/Regulation E. Regulation E defines an unauthorized electronic funds transfer as a transfer “initiated by a person other than the consumer without actual authority to initiate the transfer and from which the consumer receives no benefit.” Unauthorized use under Regulation E includes fraudulent use of information from a debit card, including card number and expiration date, to initiate an electronic funds transfer.

Card network fraud liability rules

TILA/Regulation Z and the EFTA/Regulation E set a baseline maximum of consumer cardholder liability for fraudulent transactions conducted using a credit card or debit card. The effect of this public law regime is to require the card issuer to absorb all fraud liability in excess of the maximum cardholder liability allowed under law. Given the stated purposes of TILA/Regulation Z and the EFTA/Regulation E—to protect consumers—it is not surprising that these laws are not concerned with further allocation of fraud liability after shifting responsibility from the cardholder to the card issuer. The card network rules both enhance the baseline cardholder protections established by TILA/Regulation Z and by the EFTA/Regulation E and further allocate fraud liability from card issuers to merchants based on a complicated set of rules that vary based on the type of transaction at issue. The card networks enhance the cardholder protections offered under TILA/Regulation Z and the EFTA/Regulation E through their “zero liability policies.” The card networks allocate fraud liability risk between card issuers and merchants based upon detailed dispute resolution rules, which take into account at least some element of the respective parties’ compliance with network rules designed to detect and deter attempted fraudulent transactions.

Whether the card issuer or the merchant to a particular fraudulent transaction ultimately will be liable for the fraud losses depends on if the merchant followed the payment card rules in connection with the particular transaction. There are numerous permutations of rule requirements for all manner of transaction types. One of the most significant determinants of whether the card issuer or the merchant in a particular transaction will be responsible for fraud is whether the transaction is a face-to-face transaction (a “card-present transaction”) or a transaction conducted over the Internet, by mail, or by telephone (a “card-not-present transaction”).

If one distills the standard requirements across the card networks to their essence, it is generally true that a merchant engaging in a card-not-present transaction may only successfully overcome a cardholder/card issuer allegation that the transaction was the result of fraud if the merchant 1) performed an address
verification at the time the transaction was authorized (that is, verified that the person conducting the transaction could validate the billing address associated with the payment card being used); 2) delivered the purchased merchandise to an address that matches the address validated through the address verification; and 3) obtained proof that the purchased goods were delivered to the verified address. If the merchant cannot satisfy these requirements, the card network rules typically shift fraud liability from the card issuer to the merchant. Contrast this to the card-present transaction environment, where a merchant may successfully defend a transaction disputed by the cardholder or card issuer as fraudulent by demonstrating that the card was present at the point of sale and by producing a signed transaction receipt. In the event of such a successful defense, the card issuer typically will be held accountable for the fraud losses.

**Do current fraud liability allocation rules create incentives that minimize systemwide fraud losses?**

A shorthand way to look at default liability allocation under public law and private rules of the payment card schemes is as follows: 1) Consumers rarely bear meaningful liability for fraudulent transactions unless they benefited from the fraud; 2) issuers typically bear liability for fraud losses perpetrated in card-present transactions; and 3) merchants generally bear liability for fraud losses perpetrated in card-not-present transactions. Taking a systemwide approach to fraud in card-based payment systems, the natural question that follows from the current status quo is whether the rules for fraud liability allocation result in efficient outcomes: That is, are the parties to each payment card transaction vested with appropriate incentives in the form of fraud liability risk to encourage each to take reasonable steps to minimize fraud losses viewed from the perspective of the payment system as a whole?

**Cardholder liability for fraudulent transactions**

There is little doubt that cardholders’ carelessness in protecting their own card information contributes to the incidence of payment card fraud. A recent study commissioned by Canada’s Interac Association found that 60 percent of Canadians do not shield their PIN entry at automated teller machines (ATMs) or point-of-sale terminals when they believe no one is watching them and that 37 percent do not shield their PIN entry even when they believe someone can see them entering it. The extent to which cardholders are regularly negligent in protecting their own card information from potential fraudsters is debatable. On the one hand, cardholders surely do not wish to invite fraud. On the other hand, while cardholders may not be aware of the nuanced differences in fraud liability protections available under public laws and private rules, it would be difficult for cardholders not to be aware of their protections under the zero liability policies prominently and repeatedly promoted by the card networks.

Assuming most consumers understand, at least in some abstract sense, that they are protected from liability for fraud losses regardless of their level of diligence in safeguarding their own information, one wonders whether a greater deductible on the first-dollar insurance coverage mandated by the card networks through zero liability policies would reduce the incidence of fraud by encouraging appropriate risk-avoiding behavior. As it currently stands, the major card networks’ zero liability policies (and even the very low deductibles payable by cardholders under public law) leave in place a significant risk of moral hazard that almost certainly, at least at the margins, contributes to overall systemwide fraud losses.

Notwithstanding what appears to be somewhat low-hanging fruit in the effort to achieve systemwide fraud reduction, there are two significant challenges—both likely insurmountable—that make increasing cardholder liability highly unlikely regardless of the efficiency in the outcome it may engender. The first challenge is the increasing trend among legislators and regulators to enact payment-system-related public laws that offer greater consumer protection regardless of the efficiency of the fraud-related outcomes these laws may create. A reversal of this trend among legislators, in particular, is unlikely given increased public attention on consumer protections in payment systems.

The second challenge is the need, critical to broad-based user adoption and acceptance of any payment system, for the users to have confidence in the system’s security and safety. Card network operators are constantly searching for ways to induce greater cardholder confidence in the security of making card-based payments—which they hope will result in a correlative increase in transaction volume across the payment system. Designing a card-based payment system that increases consumer liability for fraudulent transactions would likely undermine confidence in the system overall and result in reduced transaction volume—the opposite of the desired effect. Given these counterincentives among those who promulgate the applicable public laws and private rules, increased cardholder liability is likely not a viable option for improving the overall efficiency of fraud liability allocation rules.
Liability for fraudulent transactions: Card issuer versus merchant

If increasing cardholder liability is an improbable outcome of any fraud-reducing reforms to card payment systems at the level of either public law or private rules, then we are left to consider whether adjustments to the allocation of fraud liability between card issuers and merchants under current card network rules might have a desirable effect in reducing systemwide fraud losses. As described previously, the card issuers generally bear fraud liability in card-present transactions and merchants generally bear fraud liability in card-not-present transactions.

In the card-present context, existing card network rules may provide inadequate incentives for merchants to take efforts to detect and deter fraudulent transactions. Generally, so long as the presented card is swiped through the point-of-sale terminal and a signature is obtained on the transaction receipt, the merchant will not bear the loss if the transaction is subsequently challenged as fraudulent. Consequently, the marginal economic benefit to merchants of deploying additional fraud prevention measures, even if effective measures are made available by card issuers and card networks, may well not justify the costs to the merchant of implementation because the merchant stands to gain little. Fraud detection measures in traditional brick-and-mortar sales channels today include the examination of the card for evidence of tampering and a comparison of the signature on the transaction receipt to the signature on the back of the card (although many merchants’ employees do not even glance at the card presented for payment).

In contrast, in the card-not-present environment, existing card network rules may create disincentives for card issuers to support and induce their cardholders to participate in fraud prevention efforts. Nowhere is this more evident than in the surprisingly low adoption of card networks’ payer authentication programs. Visa and MasterCard have each developed and actively promoted services designed to assist Internet merchants in authenticating payers—for Visa the Verified by Visa program and for MasterCard the MasterCard SecureCode program. Under both programs, a pre-enrolled cardholder conducting a card-not-present transaction at a participating merchant is asked to provide an authenticating password in a secure pop-up window or frame linked to the card issuer. The pop-up window or frame in which the cardholder is asked to provide the password displays a phrase or image preselected by the cardholder so that the cardholder can validate that the pop-up or frame is linked to the card issuer. This bidirectional layer of additional authentication not only deters fraud, but card network rules provide that it also shifts fraud liability risk from the merchant to the card issuer for the verified transaction.

One might think merchants would eagerly adopt these additional security measures and embrace the attendant liability shift to the card issuer for Internet transactions. However, online merchants that have attempted to require customers to enroll in such programs have invoked the ire of their customers. Card issuers have little incentive to expend resources or risk cardholder backlash by requiring participation in such programs given that the benefit would accrue primarily to the merchant, with the added offense of shifting transaction fraud liability to the issuer. In other words, card network rules appear to create the same dilemma of moral hazard in allocating fraud losses between card issuers and merchants in both card-not-present and card-present transactions as is created by public laws and private rules that insulate cardholders from fraud liability.

Conclusion

Empirical evaluation suggests that current public law regimes and private card network rules may fail to create appropriate incentives for cardholders, merchants (in card-present transactions), and card issuers (in card-not-present transactions) to adopt fraud-reducing practices. These rules may also discourage fraud-avoiding behavior in certain circumstances because of the associated costs and efforts involved and the limited benefit to be gained by the party undertaking those costs and efforts. This is not to say the current architecture of public laws and private rules is fundamentally flawed or in need of reworking from the ground up. As Robert Ballen and Thomas Fox have argued, the current system in which public law and private rulemaking collaborate to create fraud liability rules is capable of functioning effectively to achieve efficiency in payment systems. However, it may be time to reevaluate the incentives created by current card network rules in allocating fraud liability among transaction participants to better align risks with the parties that are able to make efficient decisions regarding how to mitigate them. Increasing cardholder liability is likely not on the table for consideration, but efficiency gains in terms of reduced systemwide fraud losses may well be possible through relatively minor adjustments to the allocation of liability between merchants and card issuers.
NOTES


5Card payment fraud can be perpetrated in person, if the fraudster has obtained the actual payment card, or over the Internet or via mail or telephone order, or if the fraudster possesses the victim’s name, card account number, expiration date, and card identification number (CID)—also called card verification value (CVV2) or card verification code (CVV2), depending on the card scheme at issue. Many online retailers will accept, and card issuers will approve, transactions in which significantly less information is provided through the online payment channel.

6One author has suggested that the victims of identity theft spend an average of 40 hours resolving fraudulent transactions and other issues relating to the identity theft. See Erin Fonté, 2007, “Who should pay the price for identity theft?,” Federal Lawyer, September, pp. 24–25.

7The Truth in Lending Act, which is contained in Title I of the Consumer Credit Protection Act, as amended (15 U.S.C. § 1601 et seq.), was enacted by Congress in 1968 as a consumer protection measure requiring clear disclosure of key terms and costs of lending arrangements. The Federal Reserve Board has promulgated Regulation Z to implement TILA pursuant to authority granted under 15 U.S.C. § 1607. The Electronic Fund Transfer Act (15 U.S.C. § 1693 et seq.) was enacted by Congress in 1978 to establish rights, liabilities, and responsibilities of consumers who use and financial institutions that offer electronic fund transfer services. The Federal Reserve Board has promulgated Regulation E to implement the EFTA pursuant to authority granted under 15 U.S.C. § 1693b.

8This historical trend has been threatened as of late. During 2007 and 2008, Congress became much more active in proposing and passing identity protection bills. See, for example, Credit Cardholders’ Bill of Rights Act of 2008 (H. R. 5244) and Credit Card Reform Act of 2008 (S. 2753).

9Like Congress, the Federal Reserve Board was much more active in addressing payment market developments during 2008 than it had been historically. For example, on May 19, 2008, the Federal Reserve Board proposed an uncharacteristically sweeping set of amendments to Regulation Z. See Board of Governors of the Federal Reserve System, 2008, “12 C.F.R. Part 226, [Regulation Z; Docket No. R–1286], Truth in Lending; proposed rule,” Federal Register, Vol. 73, No. 97, May 19, pp. 28866–28901. Much of this activity likely stems from the push the Federal Reserve Board is feeling from Congress. The rapid-fire succession of consumer protections bills from Congress appears to have served as a sort of notice to the Federal Reserve Board to regulate or get out of the way.


12See id.

132 C.F.R. § 226.12 n.22.


15See 12 C.F.R. § 205.6(b)(1).

16See 12 C.F.R. § 205.6(b)(2).

17See 12 C.F.R. § 205.6(b)(3).


1912 C.F.R. § 205.2(m).

20As mentioned in the subsection titled payment systems fraud generally versus signature-based card fraud, the present discussion is limited to signature-based debit and credit cards and card networks—meaning that the card network rules considered are those of Visa U.S.A. Inc., MasterCard International Inc., American Express Travel Related Services Company Inc., and Discover Financial Services.

21Readers may note that TILA/Regulation Z apply the $50 liability cap to all cardholders, not just consumers, while the EFTA/Regulation E apply the limitations on liability only to consumer cardholders. While this is a meaningful distinction, we will assume for purposes of the present discussion that the victimized cardholder is a consumer, whether the card at issue is a credit card or a debit card.

22According to 15 U.S.C. 1601, “it is the purpose of [TILA] ... to protect the consumer against inaccurate and unfair credit billing and credit card practices.” According to 15 U.S.C. 1693, “the primary objective of [the EFTA] ... is the provision of individual consumer rights.”

23Regulation Z expressly provides that an agreement between a cardholder and the card issuer may impose lesser liability on the cardholder than is provided for under Regulation Z. See 12 C.F.R. § 226.12(b)(4). Similarly, Regulation E acknowledges that a cardholder and card issuer may agree to a lower cardholder’s liability limit than the Regulation E default. See 12 C.F.R. § 205.6(b)(6).

24Each of Visa, MasterCard, American Express, and Discover has enacted some form of zero liability policy. The ultimate effect is that, except in very limited circumstances, a card issuer is required to assume, on behalf of its cardholders, even the amount of fraud liability permitted to be passed on to the cardholder under applicable public laws.

24Commentators have noted that consumers are unaware of the different regulatory protections that apply based on the source of funding supporting a payment card transaction. See, for example, Marianne Crowe, Scott Schuh, and Joanna Stavins, 2006, “Consumer behavior and payment choice: A conference summary,” Public Policy Discussion Papers, Federal Reserve Bank of Boston, discussion paper, No. 06-1, available at www.bos.frb.org/economic/ppo/2006/ppo0601.pdf, and Furletti and Smith (2005).


26The same argument applies, albeit to a slightly lesser degree, to the very low deductibles payable by cardholders in connection with fraud loss insurance mandated by TILA/Regulation Z and the EFTA/Regulation E. In order to truly cause cardholders to take note of their liability exposure and adjust their behavior appropriately, the cardholder deductible would likely need to rise to a level exceeding the current public law maximums.

27Moral hazard has been defined as “the tendency for the insurance plans to encourage behavior that increases the risk of insured loss” by Allard E. Dembe and Leslie I. Boden, 2000, “Moral hazard: A question of morality?,” New Solutions, Vol. 10, No. 3, pp. 257–279. That definition is consistent with the use of the term for this discussion. If a participant in a given payment system has no risk of loss due to fraudulent transactions, that participant may have little incentive to take actions, even of the simplest nature, to avoid or reduce the likelihood of fraud occurring.

28Without broadening the discussion to politics in general, recent proposals in Congress to enhance consumer protection laws that would increase substantially the costs to lenders of extending credit—as well as recent amendments to Regulation Z proposed by the Federal Reserve Board to do the same—support this proposition. See note 8.


31See Federal Reserve Bank of Philadelphia, Payment Cards Center, 2003, “After the hype: E-commerce payments grow up,” discussion paper, No. 03-12, available at www.philadelphiafed.org/pcc/conferences/2003/eCommerce_062003.pdf (see, in particular, the summary of the presentation by Steven W. Klebe, titled “Online fraud: The stakes are rising”).

32See id.

33See, for example, Josh Leyden, 2008, “Net shoppers bullied into being verified by Visa,” The Register, August 7, available at www.theregister.co.uk/2008/08/07/verified_by_visa_compulsion/.

Vulnerabilities in first-generation RFID-enabled credit cards

Thomas S. Heydt-Benjamin, Daniel V. Bailey, Kevin Fu, Ari Juels, and Tom O’Hare

Introduction

An increasing number of credit cards now contain a tiny wireless computer chip and antenna based on RFID (radio frequency identification) and contactless smart card technology.1 The RFID-enabled credit cards permit contactless payments that are fast, easy, and often more reliable than magnetic stripe card transactions, and only physical proximity (rather than contact) is required between this type of credit card and the reader. An estimated 20 million RFID-enabled credit cards and 150,000 vendor readers are already deployed in the U.S. (Bray, 2006). According to Visa USA, “This has been the fastest acceptance of new payment technology in the history of the industry” (Bray, 2006).

The conveniences of RFID-enabled credit cards also lead to new risks for security and privacy. Traditional (magnetic stripe) credit cards require visual access or direct physical contact for retrieving information, such as the cardholder’s name and the credit card number. By contrast, RFID-enabled credit cards make these and other sensitive pieces of data available using a small radio transponder that is energized and interrogated by a reader.

Experimental results

Although RFID-enabled credit cards are widely reported to use sophisticated cryptography,2 our experiments found several surprising vulnerabilities in every system we examined. We collected two commercial readers from two independent manufacturers and approximately 20 RFID-enabled credit cards issued in the last year from three major payment associations and several issuing banks in the U.S. We were unable to locate public documentation on the proprietary commands used by RFID-enabled credit cards. Thus, we reverse-engineered the protocols and constructed inexpensive devices that emulate both the credit cards and readers. The experiments indicate that all the cards are susceptible to live relay attacks (in which an attacker relays verbatim a message from the sender to a valid receiver of the message), all the cards are susceptible to disclosure of personal information, and many of the cards are susceptible to various types of replay attacks (a form of network attack in which a valid data transmission is maliciously or fraudulently repeated or delayed). In addition, we successfully completed a proof-of-concept cross-contamination attack.

Given the size and diversity of our sample set, we believe that our results reflect the current state of deployed RFID-enabled credit cards; however, card issuers continue to innovate and will likely add new security features. Our findings are not necessarily exhaustive, and there may exist cards that use security mechanisms beyond what we have observed.

Background

In this section, we provide some background on the current state and standards of RFID technology and its deployment throughout the United States.

Thomas S. Heydt-Benjamin is a former graduate student at the University of Massachusetts Amherst. Daniel V. Bailey is a senior research scientist at RSA Laboratories in Bedford, Massachusetts. Kevin Fu is an assistant professor in the Department of Computer Science at the University of Massachusetts Amherst. Ari Juels is a chief scientist and director at RSA Laboratories. Tom O’Hare is an employee of Innealta Inc. in Salem, Massachusetts. The authors thank Russell Silva for his assistance in implementing Linux drivers for RFID devices as part of his undergraduate research project at the University of Massachusetts Amherst. They thank Robert Jackson and Prashant Shenoy for sharing their laboratory equipment. They further thank the anonymous reviewers, Simson Garfinkel, Yoshi Kohno, David Molnar, and Adam Stubblefield for reviewing earlier manuscripts. This research was supported in part by grants from the National Science Foundation (award Nos. CNS-052072 and CNS-0627529).
Scale of current deployment
Several large chain stores in the U.S. have deployed many thousands of RFID readers for credit cards: CVS Pharmacies (all 5,300 locations), McDonald’s (12,000 of 13,700 locations), the Regal Entertainment Group of movie theaters, and several other large vendors (Koper, 2006; and O’Connor, 2006). Reports estimate that 20 million to 55 million RFID-enabled credit cards are in circulation, which is 5 percent to 14 percent of all credit cards (Averkamp, 2005; Bray, 2006; and Koper, 2006). In addition to traditional payment contexts, RFID-enabled credit cards are becoming accepted in other contexts such as public transportation (Heydt-Benjamin, Chae, et al., 2006). The New York City subway (Metropolitan Transit Authority, 2006) recently started a trial of 30 stations accepting an estimated 100,000 RFID-enabled credit cards (SourceMedia Inc., 2006). A participant in this trial uses her credit card as a transit ticket as well as a credit card in place of the traditional magnetic-stripe-based dedicated subway tickets.

Integration of radio frequency technology into existing credit card infrastructure
In a typical deployment, an RFID-enabled credit card reader is attached to a traditional cash register. Each reader continually broadcasts a radio signal to which RFID-enabled credit cards can respond. The RFID-enabled payment cards that we examined seem to have been designed specifically for easy integration into the existing payment authorization infrastructure. For instance, even though no magnetic stripes are read during an RF transaction, the RFID-enabled credit card readers that we examined reformat the received RFID data into “Track 1 Data” and “Track 2 Data” before passing them along to point-of-sale (POS) terminals. In other words, data are presented to the charge-processing network in the same format regardless of whether the credit card reader received the information from an RF transaction or a traditional swipe of a magnetic stripe.

Our work focuses on the first step in a long chain of system interactions: card presentation. When considering the potential impact of the vulnerabilities we have observed in RFID-enabled card presentation, one must take into account the expertise credit card issuers have gained in detecting fraudulent transactions by tracking patterns of behavior (Dougherty, 2000). While detecting fraud is an effective defense against many types of financial risk, it does not prevent invasion of privacy. Our study considers vulnerabilities to privacy that today’s antifraud methods do not prevent.

Communications protocol used by RFID-enabled credit cards
All of the credit cards we tested use a communications protocol specified by the International Organization for Standardization (ISO) in a series of documents titled ISO 14443-1 through 14443-4. Our experiments indicate that the cards use the B version of this protocol, with an additional proprietary communications layer carried over ISO layer 4.

Related work
RFID-enabled credit cards share many of the challenges and approaches for security and privacy as other RFID-based authentication and identification systems. We discuss some of these here.

RFID authentication and cloning
Many types of RFID tags merely emit static identifiers, making them easy to clone. These tags are sometimes used in inappropriate contexts such as building access control. Westhues (2005) has demonstrated a simple, inexpensive device that can skim many types of cards at a distance—even through walls—and then simulate them. (Skimming is the theft of credit card information used in an otherwise legitimate transaction.) If unclonability is a security assumption, then this is a security break.

More sophisticated tags do not emit static data, but use cryptography to emit different data during different transactions. For example, the Texas Instruments’ digital signal transponder (DST) is present in the ExxonMobil Speedpass (a keychain RFID device), and is also part of a common theft deterrent system for automobiles. These systems have been shown to be vulnerable because of faulty cryptography (Bono et al., 2005). In contrast with the RFID-enabled credit cards we examined, the DST uses cryptography to increase the difficulty of cloning, but it does not carry personally identifying information, for example, the name of its owner.

Read ranges
Industry claims around the security of RFID devices often hinge on their short read ranges. Some cautionary notes are in order, however: RFID tags do not have a single, definitive read range (Juels, 2006). While the nominal read range of an RFID tag may be quite short, a nonstandard reader or large antenna can increase the range at which an attacker can skim an RFID tag. The credit cards we examined are ISO 14443-B cards with a nominal range of 4–5 centimeters. Skimming ranges of over 20 centimeters have been demonstrated for cards of this type (Hancke, 2006),
and ranges of up to 50 centimeters are hypothesized in the literature (Kfir and Wool, 2005). Furthermore, while skimming requires that a reader power the targeted tag, an attacker performing passive eavesdropping on a session between a legitimate reader and RFID tag can potentially harvest tag data at a considerably longer range. Claims have surfaced of tests where e-passports, which rely on the same ISO standard as credit cards, were read at a distance of 30 feet (Yoshida, 2004) and detected at a distance of 20 meters (EPIC, 2005).

Our study makes no claims about the read ranges of RFID-enabled credit cards beyond the observation that characterization of these ranges is not straightforward and constitutes an important open research question.

Methodology and experiments

The following discussion highlights our methodology for testing the security of RFID-enabled credit cards against eavesdropping, skimming, and replay. A more detailed version is available in our technical report (see Heydt-Benjamin, Bailey, et al., 2006).

Eavesdropping experiments

In our eavesdropping experiments, we observed transactions between readers and cards with an oscilloscope attached to an antenna. Examination of data thus obtained demonstrated the efficacy of this simple attack, since in all transactions the cardholder’s full name and card expiration date were present in “cleartext” (that is, this information was in a form that was immediately comprehensible to a human being without additional processing, implying a lack of cryptographic protection). A majority of cards examined transmitted the credit card number in cleartext, while a minority broadcast a separate (but static) credit card number apparently reserved for wireless transactions. We provide further details in the analysis and results section.

Skimming experiments

In our most simple skimming experiment, we took a commercial RFID-enabled credit card reader and presented it with each of our experimental cards, obtaining in each case ISO 7813 (magnetic stripe style) data. Since these are the exact data normally transmitted by a POS terminal to a charge-processing network, this most naive of skimming attacks is sufficient for perpetration of certain kinds of financial fraud.

We programmed an RFID reader not intended for credit card use to emulate an RFID-enabled credit card reader. Eavesdropping on transactions between our credit card reader emulator and real RFID-enabled credit cards demonstrated that all of the RFID credit cards we tested responded to our emulator exactly as they respond to a commercial RFID-enabled credit card reader. This strongly suggests that cards do not use any secure mechanism to authenticate an authorized RFID reader before releasing sensitive information.

Replay experiments

Our credit card emulator is a microprocessor controlled device with a simple radio, permitting broadcast of arbitrary bytes over the ISO 14443-B transport layer.

We programmed our credit card emulator to expect the RFID-enabled credit card reader commands that we captured during eavesdropping experiments and then to transmit replies captured from real RFID-enabled credit cards during a skimming attack performed with the reader emulator. In our experiments, commercial readers were unable to distinguish between our emulated card and the real card upon which it was based.

Since the output from the card emulator is identical to that of the real card from which it was skimmed, a simple replay attack using this device would succeed. As noted previously, many pieces of data go into an overall transaction approval decision, including sophisticated risk-based fraud detection mechanisms on the back end. For this reason, valuable future research would include field tests in which a credit card emulator is used to perform a purchase in a retail location rather than in a laboratory.

Analysis and results

To protect the identity of our cards, we label the cards A, B, and C based on semantic equivalence classes determined by observing behavior between cards and readers. Table 1 summarizes some of the vulnerabilities of three classes of cards.

Observations of RFID-enabled credit card protocols

This section explores some of the RFID-enabled credit card protocols that are in current deployment. The analysis is based on the ISO 7813 (magnetic stripe format) data output by the serial port of RFID-enabled credit card readers when presented with different types of credit cards. Where pertinent, our analysis compares this serial output with the raw RF data from the same transactions as captured by our eavesdropping apparatus.

In keeping with a philosophy of ethical attacks research, we have redacted several pieces of information from the following subsections in part to prevent criminal misuse of our findings. The cardholder’s name and the card number have been concealed. Additionally, we have obscured the number of digits in the card number in order to obscure which observations
TABLE 1

Vulnerabilities of three classes of cards

<table>
<thead>
<tr>
<th>Card type</th>
<th>Payment association</th>
<th>Privacy invasion?</th>
<th>Relay attack? a</th>
<th>Cross-contamination?</th>
<th>Replay attack?</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>Yes</td>
<td>Yes</td>
<td>Limited b</td>
<td>Yes c</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
<td>Limited</td>
<td>Limited</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Limited</td>
</tr>
</tbody>
</table>

a Because the cards have no shielding or notion of time, all the cards are susceptible to relay.
b This attack is proven in the field, but is limited to certain merchants.
c This card admits unrestricted replay for the readers we tested, while the others induce a race condition.

Notes: This is a summary of susceptibility to various attacks for the three semantic types of cards (A, B, C) from three payment associations (1, 2, 3). A relay attack is one in which an attacker relays verbatim a message from the sender to a valid receiver of the message. A replay attack is a form of network attack in which a valid data transmission is maliciously or fraudulently repeated or delayed.

correlate with the products of specific payment associations and issuing banks.

Card A protocol

When presented (RF transaction) with any sample of a card of type A, our reader outputs serial data identical to the data contained on the magnetic stripe of the same credit card (see figure 1). When presented with the same card, the output is always the same: In the serial output there is no evidence of a counter, one-time password, or any other mechanism for prevention of replay attacks.

Card B protocol

The sample card B output in figure 2 demonstrates the presence of a counter, determined to be as such because of monotonic incrementation with successive transactions. Additionally we observe three digits that change with each transaction in no pattern that we have identified. Because of the relatively high entropy of these three digits, we consider it likely that they are the output of some cryptographic algorithm that takes the transaction counter as an input. If this is the case, then the algorithm must also take a card-specific value like a cryptographic key as an input, since we observe that different cards with the same counter value produce different codes. We speculate that these data may serve as a stand-in for the traditional card verification code (CVC).

Card C protocol

Card C’s protocol differs from card B’s in a few crucial details:

- Its unique transaction codes are eight digits instead of three;
- Its transaction counter, now located in the cardholder’s name field, displays only three digits instead of four; and
- Rather than sending the embossed card number over the air, it uses a fixed pseudonym.

See figure 3 for the sample card C output.

Analysis of RFID-enabled credit card protocols

In the following sections, we analyze the susceptibility of the card types to replay, relay, cross-contamination, and privacy/tracking attacks. Our analysis considers only the protection mechanisms of the cards and readers; we do not analyze the security of the charge-processing network (for example, the fraud detection algorithms).

Replay attacks

Replay attacks come in several flavors, depending on what data are communicated from the credit card all the way to the back-end charge-processing network. The following describe the different types of replay attacks.

- Unrestricted replay: A card that always reports the same data need be scanned only once. After that, the attacker can replay the captured data at will, and the processing network cannot detect any difference between a replay and successive transactions with a real card. Since we observed the serial output from real POS readers to always be static with respect to cards of type A, we conclude that cards of this type are susceptible to this attack.

- Replay with race condition: A card that uses a transaction counter and rolling code poses more of a challenge if the back-end processing network stores and checks counter values. In such a case, once transaction \( n \) has been accepted by the network, transactions numbered less than \( n \) should be declined if presented. However, if an adversary skims a transaction from a card and replays that transaction to the network before the legitimate user has a chance to use her card, then the charge-processing network should accept the adversary’s transactions and actually decline the legitimate ones. Although the attacker is faced with a counter synchronization problem, such challenges are far easier to defeat than the
cryptographic problems (we prefer to base our security on cryptography whenever possible).

Counter rollover: If a transaction counter is the only changing input to a code, then the number of possible codes is limited by the maximum possible transaction counter value. There are then two cases. In one case, the counter is permitted to roll over, repeating from the beginning, thus also repeating the codes from the beginning. In the other case, the card refuses to engage in additional transactions after the counter is exhausted.

In the first case, an adversary that has sufficient time in proximity to a card can build a database of all possible counter values and their corresponding codes, and therefore can mimic all possible behavior of the target card. Cards of type B are susceptible to this attack.

In the second case, a denial-of-service attack can be perpetrated against the card if the attacker has sufficient time in proximity to exhaust the counter by repeated skimming. Our experiments determined that cards of type C exhibit this behavior.

**Relay attack**

Even with a hypothetical card that combines a challenge-response protocol with a transaction counter (a case not examined here), the relay attack may still succeed (Hancke, 2005). In an example of a relay attack, the adversary consists of a **mole** and a **proxy** that perform a purchase at an innocent user’s expense. The mole possesses a clandestine **credit card reader emulator** with a (non-RFID) radio link to the
proxy’s clandestine credit card emulator. The mole sits down or stands next to the user, and the mole’s device rapidly discovers the user’s credit card. The proxy receiving this relayed signal approaches the POS terminal and initiates a purchase. The proxy presents his credit card emulator to the POS terminal. The emulator receives commands from the POS terminal and relays them to the mole’s device, which transmits the commands to the user’s credit card. The responses from the user’s card are likewise relayed through the mole’s device and are broadcast from the proxy’s emulator to the POS terminal. The purchase should succeed, and the cost will be charged to the user. Observe that even with application-layer challenge-response or transaction-counter protocols, this attack will still succeed, as protocol messages will simply be relayed between the card and reader.

Cross-contamination attack

To analyze the feasibility of a cross-contamination attack, we took a credit card of type A, placed it in a sealed envelope, and performed a “Johnny Carson attack,” by reading the card through the envelope using our custom programmed TI s4100 reader. We combined the data thus obtained with address and telephone information looked up in the telephone directory given the cardholder’s name transmitted through the envelope (for postal mail, the attacker already knows the cardholder’s address!). Using only this information we placed an online purchase for electronic parts from one of our major research parts suppliers. Our purchase was successful, and we conclude that the cross-contamination attack is effective for cards of type A and merchants that do not require a CVC.

Privacy invasion and tracking

Our eavesdropping transcripts show that personally identifying information is broadcast in cleartext by every RFID-enabled credit card we have examined. This must be considered a privacy vulnerability in that automated full identification of a person carrying an RFID-enabled credit card is easily demonstrated in the lab, and should be feasible in the field. This vulnerability is exacerbated by an adversary who could use the full identity disclosure of the RFID-enabled credit card to build up a database of associated pseudonyms based on other RFID tags with a longer read range that a user may commonly carry.

In addition, the transaction counter found in some of the cards could be exploited by a vendor: By storing the transaction counter, a retailer could tell how often the card was used to purchase goods from others. Those heavily using their cards might be targeted for specific advertising, for instance.

Countermeasures

In addition to fraud detection to limit financial risk, several other countermeasures could significantly reduce risk of fraud and invasion of privacy. We discuss some of these countermeasures here.

Shielding and blocking

One countermeasure to some cases of skimming and relay attacks is to ensure that credit cards are unreadable when not in use. A Faraday cage is a physical cover that assumes the form of a metal sheet or mesh that is opaque to certain radio waves. Consumers can today purchase Faraday cages in the form of wallets and slipcases to shield their RFID-enabled cards against unwanted scanning (DIFRwear LCC, 2006). Note that this countermeasure offers no protection when the card is in use, since a card must be removed from a shielded wallet before an RF purchase can be made. However, credit card companies ought to at least ship cards through the mail enclosed in a Faraday cage to obviate the dangers of the Johnny Carson attack.

A slightly more sophisticated approach to preventing attack against dormant RFID devices is to disrupt ambient RFID communication. Blocker tags (Juels, Rivest, and Szyllo, 2003) and the RFID Guardian (Rieback et al., 2006) are two examples of devices that can selectively disrupt RFID communications to offer tag owners improved access control.

Signaling cardholder’s intent

As an alternative approach to protections such as the Faraday cage, the credit cards themselves could be modified to activate only after indication of user intent. A simple push button would serve this purpose (Selker, 2003), but more sophisticated sensors might serve the same purpose, such as light sensors that render cards inactive in the dark, heat sensors that detect the proximity of the human hand, motion sensors that detect a telltale “tap-and-go” trajectory. Ultimately, credit card functionality will see incorporation into higher-powered consumer devices, such as near-field-communication-ready (NFC-ready) mobile phones, and will benefit from the security protections of these host devices, such as biometric sensors and increased computational capacity (Carey, 2006).

Better cryptography

Contactless smart cards capable of robust cryptography have long been available. These techniques have already been applied to payment cards in the EMV (EuroPay, MasterCard, Visa) standards, detailed in the next section. If personally identifiable data can only be decrypted by authorized readers, then the danger of many of the privacy invasion attacks discussed
here are obviated. Anecdotal accounts suggest payment associations are moving to improve the on-chip cryptographic features of these cards, including challenge-response protocols, to further frustrate replay attacks.

**Discussion**

As time goes on and technology costs decrease, we can expect issuers to provide more effective cryptographic protocols. Well-established methods to thwart these attacks already exist, and issuers may in fact already be implementing these defenses. But even today, in most cases an attacker has easier avenues to exploit than RF-based attacks to perpetrate financial fraud. For instance, simple cloning of cards is often not sufficient to commit fraud. There are many backend fraud detection measures in place to help thwart fraudulent use of card information. Nevertheless, privacy vulnerabilities should be addressed wherever they are found; privacy invasion may lead to financial fraud, but preventing financial fraud is not the only reason to protect privacy.

**Comparison with other types of fraud**

It is hard to directly compare the security of traditional magnetic stripe cards and RFID-enabled cards. RFID-enabled cards are only more secure than their traditional counterparts against certain kinds of attacks. For example, some traditional card reading mechanisms, such as taking a physical carbon copy of the face of the card, leave a physical image of the card in the hands of a possibly adversarial merchant or clerk. In fact, the use of a magnetic stripe generally means handing one’s card to a clerk who may have nefarious intent. By contrast, an RF transaction leaves behind no physical carbon copy; in fact, the card never leaves the cardholder’s hands. Certainly, the effort required to obtain an RF copy of the transaction is greater in this case.

Additionally some RFID-enabled cards include a unique code for each transaction replacing the static data in a magnetic stripe. This mechanism protects against some kinds of attacks, but creates opportunities for new types of attacks that cannot be easily addressed by traditional fraud control (such as cardholder tracking attacks).

Perhaps the most important difference between RFID-enabled cards and traditional cards is the difference in the cardholder’s control. Whereas a traditional magnetic stripe reveals one’s name and card number only when the artifact is physically handed to a merchant, an RFID-enabled card is in some sense “always on.” The card can be scanned and privacy can be compromised remotely without the knowledge or consent of the cardholder.

**Comparison with other electronic cards**

The relationship between the cards we examined and the EMV series of standards is unclear (EMVCo LLC, 2004). Certainly in Europe, EMV techniques such as the UK’s “Chip and PIN” (personal identification number) are seeing wide deployment and analysis. But based on our observations, the protocols used by the U.S. contactless cards do not appear in the EMV standards.

It is not clear to us why the U.S. payment associations have chosen to develop new protocols, with significant vulnerabilities, rather than use the more secure protocols that have already been deployed in Europe. We can surmise that this choice was motivated by the prevalence of online readers in the U.S. (some of the expense of supporting the EMV standards has to do with support for off-line operation) and a focus on contactless operation (whereas most of Europe’s cards are contact-based).

**Policy and regulation**

Several state legislatures have recently considered bills on RFID. For instance, California Governor Arnold Schwarzenegger recently vetoed his state’s bill SB 768, which would have required interim protections for RFID cards, especially cards carrying personally identifiable information, and a process for figuring out long-term protections (Ferguson, 2006; and Molnar, 2006). The information made available by the cards, including the cardholder’s name and card number are called personally identifiable information (PII) in the parlance of that bill. If signed into law, ID cards issued by the state government carrying PII would have been required to implement mutual authentication and encryption to release the data. While credit cards are not state ID cards, as time goes on we can expect more RFID-related legislation like California’s SB 768 to be introduced. Indeed, U.S. Senator Charles Schumer (D–NY) recently announced his intent to increase federal regulation of RFID-enabled credit cards (Chan, 2006).

Beyond regulation, it is an important open question as to how best to offer incentives for all custodians of personal data to take adequate precautions. Risk management is critical to the financial industry. However, as researchers and providers of risk management, we have yet to find a satisfying definition of privacy. How do we quantify user privacy when different users place a different value on privacy? In hard figures, how does this value affect the bottom line of businesses that are custodians of personal data?
Conclusion

Despite the millions of RFID-enabled payment cards already in circulation, and the large investment required for their manufacture, personalization, and distribution, all the cards we examined are susceptible to privacy invasion and relay attacks. Some cards may be skimmed once and replayed at will, while others pose a modest additional synchronization burden to the attacker. After reverse-engineering the secret protocols between RFID-enabled credit cards and readers, we were able to build a device capable of mounting several advanced replay attacks under laboratory conditions. While absolute security and privacy in a contactless card form factor may be impossible to achieve, we hope that the next generation of RFID-enabled payment systems will protect against the vulnerabilities that our study identifies.

NOTES

1This article was originally published as Heydt-Benjamin, Bailey, et al. (2008). The full version of this paper appears as a University of Massachusetts Amherst technical report (Heydt-Benjamin, Bailey, et al., 2006). See www.rfid-cusp.org for the latest version.


REFERENCES


Molnar, D., 2006, personal communication.


