ActivPass: Your Daily Activity is Your Password

Sourav Kumar Dandapat
IIT Kharagpur, India

Swadhin Pradhan
UTAustin, USA
*

Bivas Mitra
IIT Kharagpur, India

Romit Roy Choudhury
UIUC, USA

Niloy Ganguly
IIT Kharagpur, India

ABSTRACT

This paper explores the feasibility of automatically extracting passwords from a user's daily activity logs, such as her Facebook activity, phone activity etc. As an example, a smartphone might ask the user: "*Today morning from whom did you receive an SMS ?*" In this paper, we observe that infrequent activities (i.e., outliers) can be memorable and unpredictable. Building on this observation, we have developed an end to end system $ActivPass$ and experimented with 70 users. With activity logs from Facebook, browsing history, call logs, and SMSs, the system achieves 95% success (authenticates legitimate users) and is compromised in 5.5% cases (authenticates impostors). While this level of security is obviously inadequate for serious authentication systems, certain practices such as password sharing can immediately be thwarted from the dynamic nature of passwords. With security improvements in the future, activity-based authentication could fill in for the inadequacies in today's password-based systems.

Author Keywords

Activity-based password; Password sharing; Outliers; Dynamic authentication;

ACM Classification Keywords

K.4.4. Computers and Society: Security

INTRODUCTION

> *All human beings have three lives: public, private, and secret* - Gabriel Garcia Marquez

Passwords have been the canonical method of authenticating a user's identity. It has been immensely successful over the last several decades, allowing an effective balance of security and simplicity. Unfortunately, passwords are failing to scale to the changing landscape of computing. For instance: (1) With an exploding number of apps and online services, the burden of remembering site-specific passwords is almost prohibitive. Using common passwords across sites alleviates the burden but at the cost of diminished security. (2) With cloud based services, such as Netflix, users are now able to share passwords among friends – password based authentication is not fundamentally designed to thwart such behaviors. (3) The possibility of password getting stolen is increasing alarmingly, note the recent Gmail password outage[1].

In general, the need for authentication is emerging in a wide range of contexts and various solutions are necessary each operating at different points on the trade-off spectrum. This paper develops an authentication scheme that is less secure than passwords, but simpler to use and resistant to sharing. Our core idea is to observe a user's activities from the recent past and extract questions from them that, ideally, only the user can answer but others cannot. Example questions could be *"from whom did you get the first call today morning?"* or a multiple-choice format that does not require much typing – *"which news site did you NOT browse today morning: CNN, NYT, Slashdot, Wired"*. These questions will change for every instance of authentication, disallowing a single breach to cause a permanent damage. Finally, the user would be able to configure the parameters of the system, including the number of questions for successful authentication, multiple-choice or not, and of course, permissions to the activity logs. Given that today's users perform various activities jointly with their computing devices, we believe that adequate "secrets" can be extracted, enabling this alternative form of authentication.

A natural question might be *where is such activity-based authentication applicable?* While this paper is inherently an exploration of the feasibility of this authentication mechanism, we believe password sharing in Netflix or HBO like services is an area of application. Our proposal can potentially reduce such sharing — even if Alice has shared her Netflix password once with Bob, she may not be willing to share her personal activity information every time Bob uses the password. In another setting, activity-based authentication could also serve as an alternative to hint questions. When a user forgets her password, she could be authenticated through a series of activity-based questions instead of regular hint questions.

*Most of the work has been done during his MS in Indian Institute of Technology Kharagpur.

[1] http://www.ibtimes.com/5-million-gmail-usernames-passwords-hacked-posted-russian-bitcoin-forum-report-1684368

CHI 2015, April 18 - 23 2015, Seoul, Republic of Korea
Copyright 2015 ACM 978-1-4503-3145-6/15/04...$15.00
http://dx.doi.org/10.1145/2702123.2702457

This paper employs the core idea that outliers in the user's activities (rare activities) offer opportunities for generating passwords. Intuitively, outlier events are easy to remember and difficult to guess. We establish this intuition through a sequence of motivational experiments, and use the lessons to develop *ActivPass*, an automatic system that identifies activity outliers and carefully generates (textual and multiple choice) questions from them. Textual passwords offer better security while multiple choices are easier to answer with just a click — the user configures the system based on her preferences. Activity information are sourced from 3 categories, namely, (1) smartphone call logs and SMS senders, (2) web browser history, and (3) Facebook activities invisible to the public. Experiments are designed with 70 volunteers recruited from various segments of the university population – they answer questions generated from their own and others' activity logs. We conduct user studies to understand user experience – the results are quite promising. Many users admitted to sharing passwords, either because they "could not decline when a friend asked for it", or because they wanted to share it for only one event, "like a FIFA world cup final match". Of course, some raised questions on privacy but admitted that large companies such as Google, Facebook, Netflix anyway have access to their activities, and could already design passwords from them. Moreover, recent works are investigating architectures to scatter data and ensure protection of privacy [11]. All in all, an overwhelming majority of users "liked" the core idea, indicating they believe "it might work" and that they are comfortable with using it.

The rest of this paper expands on the above ideas beginning with related work, and followed by our core experimental set up. Then, we discuss results of few initial studies to establish the foundation of our hypothesis as motivation of this study. Next we elaborate the system design and the way of improving the question answering system. We then give a statistical overview of the data we are working on. After that we discuss the experimental results evaluated on 50 participants and the design of final $ActivPass$ system. With evaluation of final $ActivPass$ system we conclude this paper.

RELATED WORK

Authentication system can be conceptualized in a multi-dimensional space where perhaps the most important dimension is *security*, while other dimensions are usability, shareability, simplicity. Various solutions operate at different regions in this design space.

Text based password is the most popular authentication mechanism among the existing systems and is believed to be secure [3]. Many *physical biometric based authentication* systems have been proposed as this class of password is believed to be robust and secure. Examples of a few physical biometric authentication schemes are like face based authentication [5], fingerprint based authentication [5, 13], iris based authentication [10], audio based authentication [4, 15], gait based authentication [9]. *Graphical password* is a usable authentication mechanism where a predefined graphical image is shown to a user. The user requires to touch predetermined areas of the image in a particular sequence [2]. Another interesting approach of authentication mechanism is context based authentication [8, 16], which leverages the idea of individuality of users [19].

However, shareability - a new dimension of authentication is becoming important. Any public subscription based company generates revenue from users' subscriptions. If that credential is shared among many persons, then obviously service provider is going to lose revenue. Like Netflix, HBO-Go authorities curse this password sharing habit[2] leading to their revenue loss.

A number of contributions have been made by researchers in this direction. *Physical biometric* is the most promising solution that can avoid sharing. However, it is not feasible to apply *physical biometric* in all kinds of authentication scenario. Another alternative to restrict sharing is *HCI-based biometric*. In [18], authors reported a number of HCI-based biometric schemes. This work categorized the existing schemes into two classes -1) input device interaction based biometric (keystroke, mouse, haptic), and 2) software interaction based biometric (email-behavior, programming style, computer game strategy). [14] distinguished authentic user from non-authentic users based on key stroke pattern. However, authors reported even 50% false acceptance rate in key stroke pattern based authentication [17]. In [7], authors proposed techniques to identify sender of an e-mail by mining e-mail content.

One Time Password (OTP) is another potential option for solving the issue of sharing. As it changes dynamically, it would be difficult to share every time. OTP still can be shared to reduce subscription charge if users adopt the painstaking path of repeated sharing. Instead of a random string, if OTP contains a user's private information, it would become resistant to sharing. Another direction of research tries to find out secrets that users know for authentication. For example, to identify fraud in online credit card transaction, a user might have to face different security questions like geo-location, email-address, shipping-address, previous transactions [1]. In [12], authors identify that users remember various activities they perform on smartphones. In line with this, in [6], authors propose to capture users' daily events (on smart phone) that users remember to authenticate users. However, authors concentrate more on answer pattern of authentic users rather than designing questions that achieve high recall and low guessability.

$ActivPass$ builds upon these initial works to design a more fine tuned end-to-end authentication system by distinguishing potential of different activities through a thorough user study.

EXPERIMENT SETUP

In this section, we define the metrics of interest in evaluating ActivPass, details on participants, and the instructions

[2] http://www.cnet.com/news/netflix-ceo-curse-you-password-sharing-literally/
http://blogs.marketwatch.com/themargin/2013/12/31/netflix-takes-a-shot-against-shared-subscriptions-some-of-them-anyway/

provided to the participants during experiments. The experiments performed in the paper can be enumerated into the following four major stages - **a.** motivational experiment, **b.** pilot study (first stage experiment), **c.** second stage experiment, and **d.** experiment with the final system. The purpose and goal of each experimental stage is illustrated at the due place in the paper.

Metrics

Recall Rate (RR) - It is the percentage of correct answers given by a user say X when asked about her *own* activities. Let T_{own} be the total number of questions asked to X and X is able to answer p correctly then simply, $RR_X = \frac{p}{T_{own}} \times 100$.

Guessability (G) - Let, for the given user *X*, her friends [3] are shown T_X questions about *X's* activity; a particular question may be shown to *N* friends; in that case the question is counted *N* times in T_X. Let p be the number of questions out of T_X that are answered correctly by X's friend. Then, Guessability about *X* is formally defined as $G_X = \frac{p}{T_X} \times 100$.

Participant Details

To recruit participants, we had advertised via e-mail about our experiment with privacy policies and compensation to different academic institutions. The experiments have been conducted with agreed participants (students, researchers, professors, technical persons) with wide age range (from 18 years to 47 years where median age is 26); on average 30% of participants are female. Some of the volunteers participated in more than one stage. For participating in each stage of experiment, we have extended a dinner coupon to the volunteers.

Instruction to Participants

Each volunteer participated in two types of experiments - first one is about answering questions pertaining to their own activity and second one is about guessing the activities of their friends. For recall experiment users instructed to answer questions from their own activity. A user would receive a recall question in following format - "Whom did you send an SMS around 10 am today?" For guessability experiment, each volunteer was asked to explicitly identify a set of friends who were also participating in the same experiment. Let's say participant *X* identified *Y* and *Z* as her friends. In that case, *X* would receive guess questions in following format - "Whom did *Y* call at around 5pm today?", "Whom did *Z* chat with at around 3pm yesterday?"

MOTIVATIONAL EXPERIMENTS

In this section, we demonstrate a set of experiments to establish the key intuitions behind the methodology. First we show the existence of activities that users can easily recall and while others cannot guess with help of two metrics defined earlier. Then we show that deeper investigation uncovers a specific class - outlier activities. Next, we demonstrate the presence of outlier activities in our daily life.

Recall Rate and Guessability - Bounds

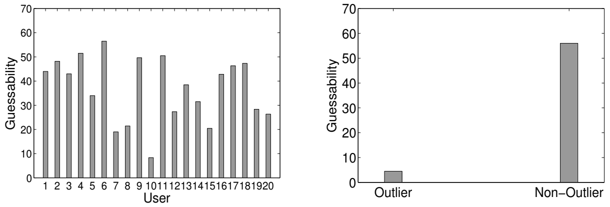
In an attempt to develop *ActivPass*, we first asked the following question: *if an extremely smart system – say humans – were to observe the activity logs and prepare questions from them, what kind of performance will it achieve?* This should serve as a reasonable estimate of upper bound on recall and a lower bound on guessability of an activity based authentication system. Insights into these bounds should shed light on how well *ActivPass* can perform. In all the experiments discussed in this section, we have considered past 7 days activities and only text-based question is prepared from activity. We have employed 20 volunteers - of which 10 are under-graduate and 6 are post-graduate students, and 4 are academic-staffs.

We designed the following three experiments:

(1) Guessability

We recruited a volunteer, say Alice, and asked her to scrutinize her own activities, and prepare 7 questions that she thinks that even her friends will not be able to guess. Then, we asked these 7 questions to $3 - 5$ individuals, selected from Alice's friends. We repeated this process for 20 volunteers altogether. Example questions that users asked were of the form: What was the last thing I ate for dinner yesterday?, What did my assistant Carol give me today? Which song did I listen to on my way back from office last night? Fig.1(a) reports the percentage guessability achieved for each volunteer. Evidently, average guess-ability is 36.8% which is (apparently) quite high for the development of any meaningful system. Nevertheless, careful investigation reveals that the set of rare (defined later) activity has very low guessability.

Fig. 1(b) shows that outlier activities consistently exhibit a marked difference with other activities with guessability rate 4.5%[4]. This encouraging result indicates that the opportunity to extract secrets from user activities indeed exists; outlier activities may play a key role. Whether it can be achieved automatically, using only a subset of collected activities, is the central question in this paper.



(a) Shows user-wise Guessability of friends' activities

(b) Guessability across different activity types

Figure 1. Guessability

(2) Recall

The second experiment is targeted to characterize the upper bound on recall. Observe that this cannot be performed the same way as above since Alice cannot be asked to identify questions that she herself remembers! Thus, we needed another individual who is a very good friend of Alice so that

[3] by friend in this context we mean class-mate/ room-mate/ lab-mate/ colleague

[4] In selected activities 35% are outlier, and 65% are non-outlier.

he/she is able to understand Alice's context, what she remembers well, her behavioral patterns, etc. Let's say this friend is Bob. We then ask Bob to look at Alice's activities of last seven days and craft 10 questions that he thinks Alice will be able to recall. Example questions that Bob crafted were of the form: Who was Alice with on a long phone call today? What song did Alice listen to during dinner last night? We repeated this process for 10 pairs of individuals. Fig. 2 shows the per-user recall rate for each of the 10 volunteers. The average rate is 89.9%. High recall rate again verifies the existence of the core opportunity – that it is possible to pick from recent activities that a user can herself recall.

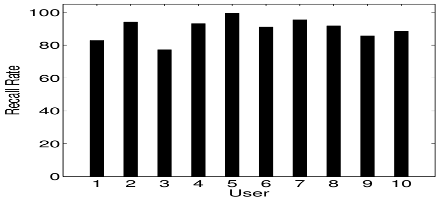


Figure 2. Shows user-wise Recall Rate of activities in recent past

(3) Satisfying Guessability and Recall

Of course, a question that is effective for guessability may not be effective for recall, and the vice versa. We need to craft questions in such a way that both the factors got satisfied. Moreover, as we will not have complete know how about user's activity, we have to prepare questions from a subset of activities that are recorded electronically. The above experiments individually point to the opportunities but do not capture these constraints for a practical system. Therefore, we perform the following experiment where we ask Bob to look into Alice's recorded activities[5] and craft questions that Alice can answer, and others cannot. The recorded activities contained browser history, Facebook activity, phone call meta-data, and SMS meta-data. We repeated this for 10 pairs of users. Fig. 3 shows the average recall and guess-ability for the crafted questions – recall rate is 87% and guess-ability is 5.3%. The results are encouraging, and indicative of the latent potential in activity based authentication.

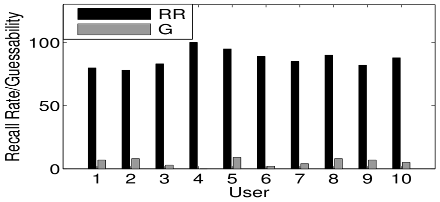
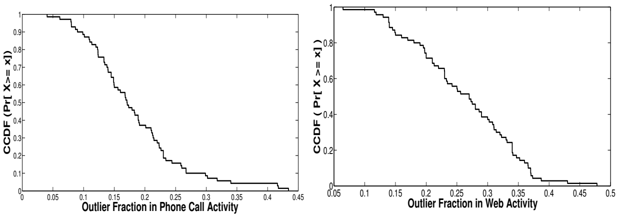


Figure 3. Shows user-wise Recall Rate and Guessability

Presence of Outlier Activities

We perform several measurements based study to establish the fact that the outliers[6] constitute a significant portion of the users daily activities which can be measured in terms of *fraction of outlier activity* performed by a particular user.

Fig. 4 plots the CCDF (Complementary Cumulative Distribution Function) of outlier activity for phone call (a user gets a call from her friend after a long time) and browsing activity (a user visited reddit.com which she normally does not visit). Interestingly, we observe that, for phone call and browsing activities, all 70 participants have outliers. Fig. 4(a) shows that almost 80% users have at least 10% outlier activity fraction in case of phone call. Fig. 4(b) illustrates that almost 100% users have around 10% or more browsing outlier activities. For SMS also we found similar pattern as call.



(a) CCDF of outlier phone call activity

(b) CCDF of outlier web links in browsing activity

Figure 4. Studies about outliers in phone call and browsing activity

SYSTEM ARCHITECTURE OF ACTIVPASS

The system in the background runs an activity listener which continuously extracts metadata from various user interactions, and organizes them into time-stamped activity logs. Example metadata are caller IDs, SMS senders, duration of calls, webpage URLS, webpage titles, Facebook profiles visited, etc. Now, when a user invokes an application, say *Netflix*, $ActivPass$ executes a *Password Generator Module (PGM)* in the background. The PGM operates on the activity logs and creates n password questions that are then presented to the user together. If the user correctly answers at least k out of n questions, $k \leq n$, the application is launched. Fig. 5 sketches the flow of operation in ActivPass.

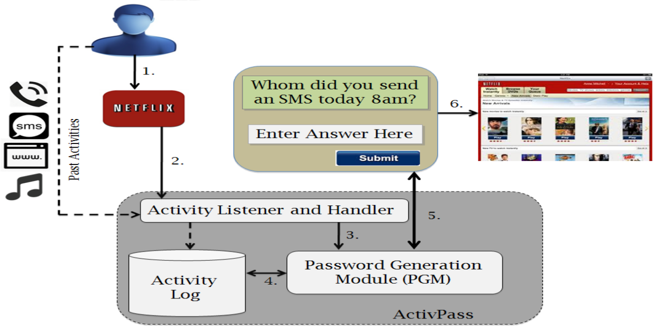

1.
NETFLIX
2.
sms
www.
Past Activities
Whom did you send an SMS today 8am?
Enter Answer Here
Submit
6.
Activity Listener and Handler
5.
3.
Activity Log
4.
Password Generation Module (PGM)
ActivPass

Figure 5. Work flow of *ActivPass* showing an instance where a user invokes the Netflix application

Password Generation Module (PGM)

Fig. 6 shows a high level flowchart of the PGM module running inside ActivPass. The Activity Handler module — installed in the form of a web browser plug-in, a Facebook app, a SMS/phone-call monitor, and an audio monitor — listens and collates all the activities from its user. PGM

[5] These activities include Facebook, sms, call, web-browsing etc. which we later use while developing the ActivPass system.

[6] An outlier activity is defined as one whose frequency of occurrence is less than 10% of the frequency of occurrence of the activity which lies on 75^{th} percentile, if we arrange the activities in order of its occurrence. This is an ad-hoc definition which we came up after observing a lot of data.

generates challenges from collected activities and provide authentication challenge to user.

Fig. 6 illustrates the operations inside the PGM. An Activity Collection module periodically draws recent activity logs and forwards them to an Activity Categorization module tasked to identify outliers. While outlier generation can be across many dimensions, we adopt a simple approach focused on only the inter-occurrence time of activities. Specifically, ActivPass computes the distribution of inter-occurrence time of each activity, and identifies those that fall in the extreme points of this distribution. Thus, if Alice visits CNN.com with some periodicity, then CNN.com is not an outlier; however, if Alice has not visited Engadget.com for a month and does so today, then it is a suitable candidate for passwords. In light of this, the categorization module computes the distribution and discards all activities that are not at the extreme of the distribution (outliers as defined in previous section). It also discards some *irrelevant* activities, such as incoming phone calls with the caller name as "Unknown Caller". The residual activities are forwarded to the Challenge Generation module which designs the questions from them. Table 1 summarizes the categories and nature of questions that the PGM ultimately presents to the user. Table 2 details the type of data collected.

Pilot Study and Data Driven (Re)Design

We subject the simple design sketch above to a pilot user study with 20 users - 10 undergraduate, 6 post-graduate students and 4 academic staffs. The goal is to gain insights on the efficacy of this basic outlier detection scheme — scenarios under which users are not able to recall, and attackers are able to guess.

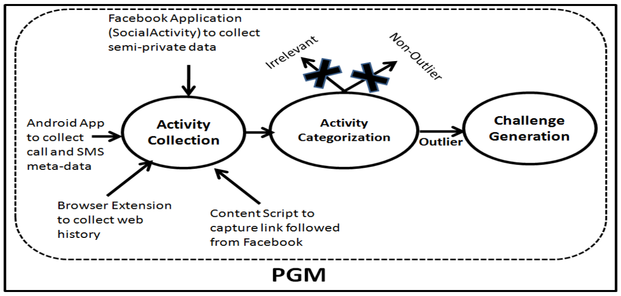

Facebook Application (SocialActivity) to collect semi-private data
Irrelevant
Non-Outlier
Android App to collect call and SMS meta-data
Activity Collection
Activity Categorization
Outlier
Challenge Generation
Browser Extension to collect web history
Content Script to capture link followed from Facebook
PGM

Figure 6. Flowchart for operations inside the Password Generation Module (PGM)

Table 1. Range of questions asked per source

	Range of questions asked
Facebook	1) Profiles visited by the user. 2) Groups the user is a member of. 3) A person with whom user had a chat.
Web	1) Titles of the web-pages visited by the user.
Call	1) A person whom the user called. 2) A person who called the user.
SMS	1) A person whom the user sent an SMS. 2) A person who sent an SMS to the user.
Audio	1) The tune/tone used by the user as an alarm. 2) The tune/tone used by the user as her ring-tone. 3) The audio files downloaded by the user.

We have gathered activities over a span of 7 days. As mentioned earlier, the activities are only meta data — we perform no content inspection. Volunteers were also given the control to turn off ActivPass when they desired to perform certain private operations. Of course, its possible that user activities were biased based on this background listener running on their devices — we decided to live with this problem for now to get a first-cut measurement. Perhaps it is worth noting, that several users mentioned being conscious in the first day or two, but forgetting about the listener after that.

Table 2. Data collection details

Source	Details of data collected
SMS	Time, Receiver/Sender Name
Call	Time, Type (incoming, outgoing), Name of other person, Duration
Audio	Title of Music added in this week, Alarm tone, Ring tone
Web	URL, Time of visit
Link visited from Facebook	URL, Time of visit
Facebook Group	Name of Private (secret and closed) groups
Facebook Pages	Name of pages created by user
Facebook Profile	Name of Facebook friends of user
Facebook Message	Time (in milliseconds from epoch), Name of other person, Msg Id, Thread Id

Running the outlier detection algorithm on the collected data, the PGM generated questions, 65 of which must be answered by each volunteer. Among them, 40 questions are from the volunteer's own activities and 25 questions are from others' activities. The questions were sourced uniformly from each of the five categories, namely web, FaceBook, phone, SMS, audio subject to the availability of enough questions in each category. Moreover, it is ensured that each user gets equal number of questions from all question formats (discussed shortly). Once volunteers answered these questions, we obtained a total of around 1300 responses — 800 of them were memory recalls and 500 were guesses to others' password questions.

Question Formats: Questions are of two formats, namely text-based and multiple choice questions (MCQ). The text-based questions are generated in a templatized manner and hints (in form of some character(s) of the answer word) may or may not be given. MCQs on the other hand are designed in a way that the outliers are camouflaged with 3 additional activities which actually never happened or happened long ago. Finally, to evaluate the design point of extreme simplicity and weak security, we also design some questions with binary (Yes/No) answers. Table 3 shows examples of different question formats.

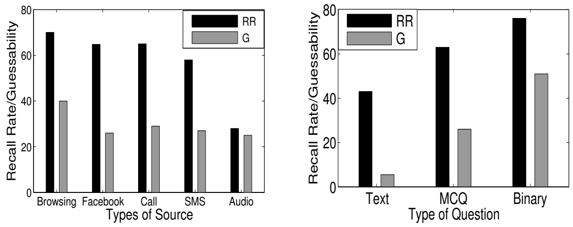
Table 3. Exemplar questions of different question formats

Question formats	Example questions asked
Binary	Have you received a call from Alice at around 10 pm on 19/09/2014?
MCQ	Please write the options of the links you visited,this week in comma separated way (Ex: A, B): A. CNN; B. BBC; C. SKY News; D. Reuters
Text	Whom did you call at around 7 pm on 17/09/2014 ? Hint: (Al*)

Observations and Results

In this section, we analyse the data collected to understand the flaw of ActivPass system design which in turn will help to improve the design and tuning the parameters of ActivPass system.

Across Activity Categories : Fig. 7(a) shows recall rate and guessability across different activity categories from which questions are generated. The recall rate percentage of web, Facebook, call and SMS related activities are 70%, 64.8%, 65%, 58% respectively, certainly lower than our expectations. Surprisingly, recall rate of audio related activities was excessively low, ≈ 28%. The excessive low recall for audio is due the fact that users load audio files in batches hence are mostly unaware about the individual songs which have been loaded. Averaging across all of them, the recall rate bordered around an unsatisfactory 61%. Fig. 7(a) also shows the guessability of activities from different activity categories. Again, volunteers were able to guess reasonably well, varying between 25% - 40%. Specifically, guessability of web, Facebook, call, SMS and audio were 40%, 26%, 29%, 27% and 25%, respectively. This higher value of guessability mainly stems from binary formats that we show shortly.



(a) (b)

Figure 7. Fig. 7(a): source-wise Recall Rate and Guessability. Fig. 7(b): Recall Rate and Guessability across different question-formats

Across Question Formats : Additional results shed more light on performance. Fig. 7(b) plots the recall rate and guessability for three question formats — (1) text-based, (2) multiple choice (MCQ), and (3) binary (yes/no). Rather surprisingly, the recall rate of even binary questions is low (≈ 76%); for MCQs, the recall rates are low too (≈ 63%). Text based questions achieve recall rates of (≈ 44%). In terms of guessability, Fig. 7(b) reports encouraging results for text based questions of around 5.5%. Also, as expected, it is around 25% for MCQ, and 51% for binary questions.

Effect of Hint : Fig. 9 shows the effect of hint, hint position and hint length in terms of recall rate and guessability. Fig. 8 shows significant difference between questions with hint and without hint in terms of recall rate. With hint, recall rate is approximately 81% while it is around 17% when there is no hint. Fig. 9(a) shows the effect of hint position. In this experiment, we have considered three different hint positions -i) at the beginning ii) at middle iii) at the end. For example if answer of some text based question is Alice the hints will look like i)A* ii) *i* iii) *e (where * means any number of characters) respectively. Hint at starting position is most effective for recall while there is no significant difference in guessability. Fig. 9(b) shows the effect of hint length. Recall rate is almost 10% higher with 2-character hint compared to 1-character hint while guessability is also 0.4% higher for 2-character hint.

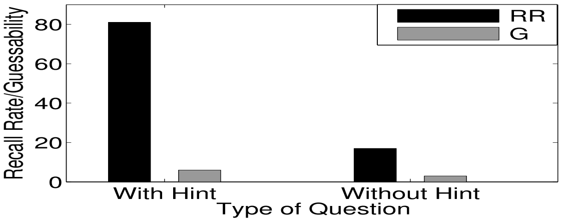
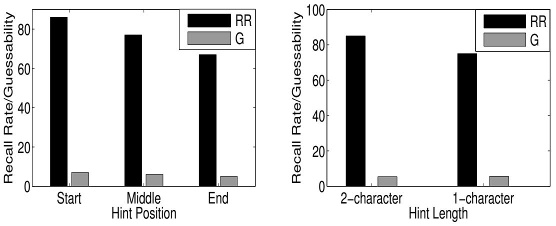


Figure 8. Recall Rate and Guessability of hint and non-hint Text questions.



(a) (b)

Figure 9. Fig. 9(a) shows effect of hint position on recall and guess. Fig. 9(b) shows effect of hint length on recall and guess

Table 5. Guessability of various source-question format combinations

Source Formats	Facebook	Web	Call	SMS	Audio
Binary	46.67	52.33	53.33	53.33	47
MCQ	27	27.33	25.67	20	22.33
Text	6.33	N/A	5.67	4.33	5.33

Effect of Staleness

Fig. 10 plots the variation of recall rate and guessability across different days of activity. It is obvious from the figure that recall rate tapers of significantly with staleness after the first three days (after that recall rate goes below 80%). However, interesting observation is that the staleness does not have effect on guessability. Guessability is independent of the activity staleness.

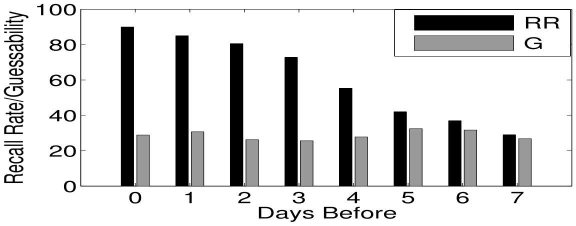


Figure 10. Impact of staleness on recall rate and guessability

Take-Away from First Stage Results

Tables 4 and 5 summarize the recall rate and guessability of all question categories from this pilot study. The results are clearly far below what can be used in a real systems. However, the following are key take-away observations that are embraced in redesigning ActivPass.

Table 4. Recall Rate for various source and question-format combinations

Formats \ Source	Facebook		Web	Call		SMS		Audio	
Binary	Chat	87	76	85		80		60	
	Group	75							
	Pages	65							
	Profile	70							
MCQ	Chat	77.5	66	Person	70	Person	80	47.5	
	Group	60							
	Pages	60		Time	75	Time	35		
	Profile	55							
Text-based	Hint	85	N/A	Hint	80	Hint	85	Hint	35
	Non Hint	17		Non Hint	20	Non Hint	15	Non Hint	10

1. *People forget their online activities quickly.* It is important to utilize very recent information for generating passwords while mere guessing is independent of staleness.
2. *Even url of web-pages are inadequate for remembering.* Several users were not able to recall whether they browsed a "lsbf.org.uk" website, but immediately responded positively when asked if they visited the "London School of Business" site. As a result, web-page titles and descriptors are needed.
3. *Hints are helpful for recall.* A hint with two character is the best although it may slightly increase the guessability.
4. *Interdependency between questions can leak information.* We observed cases where a guesser was able to break the password based on the following two questions: *When did Alice (a participant) call his friend Bob?* and *Whom did Alice call at around 5 pm today?*. Such inter-dependencies need to be handled by generating a single question per activity.
5. *With Facebook based questions, the groups and web-pages that users visit should be uncorrelated to his/her own profile.* Several "friends" were able to predict, say, that a student of MIT was visiting an alumni group of MIT Robotics.
6. *Certain category of data sources and question formats may not be used.* Questions from audio and yes/no format questions may be discarded for next round of experiment.

The first three lessons would help in improving recall while the next two helps in fixing design errors and reducing guessability. We implemented these modifications to $ActivPass$ and progressed into a fuller scale evaluation phase with 50 users.

DATA COLLECTION STATISTICS

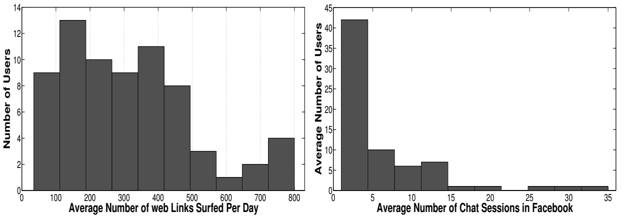
In this section, we discuss statistics of collected phone, Facebook, and web activities of 70 participants (20 for pilot study and 50 in next round). Data have been collected from a wide variety of phones like - HTC-desire, Sony-Xperia, Samsung-S2, Samsung-S3, Samsung-Duos, Samsung-Ace, Micromax-Canvas, Lenovo-A706 etc. using these applications. With due permission, we install the ActivPass listener[7] on each of their computers, laptops, and smartphones, and gather activities from them. Details of the data collected as well as the profiles of the participants are detailed below.

Profile Statistics of Participants

To understand the profiles of participants, we have conducted a survey among them. The survey revealed that the participants have possessed 2 to 10 mobiles in their lifetime. Most of the people carry multiple electronic equipments like tablet, smartphone etc.

Browsing Activity and Facebook Chat Session Statistics

Fig. 11(a) shows statistics of web browsing including link followed from Facebook. Sixty users browse in the range of 35 to 500 sites almost uniformly, while the most active 10 users browse in the range of 500 to 800 sites on an average per day. Fig. 11(b) shows statistics of Facebook chat session. Facebook chat session statistics is also quite skewed. Around 60 users participate in 1 to 10 chat sessions on an average per day while the most active 5% users have 25 to 35 chat sessions on an average per day.



(a) Histogram of average #links visited per day

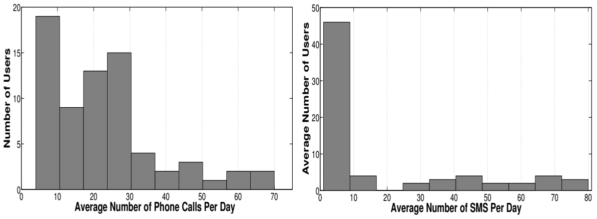
(b) Histogram of average #chat-session per day across participants

Figure 11. Browsing and Facebook activity related statistics across participants

Phone Activity

Fig. 12(a) illustrates the statistics of call data including incoming and outgoing call. It shows that number of calls per day follows an heterogeneous distribution with only 14 users making more than 30 calls, while a large portion of them doing as less as 10 calls per day. Fig. 12(b) shows statistics of SMS data including incoming and outgoing SMS. SMS statistics is clearly much more skewed than call statistics. Most of the participants do receive and send between 1 and 10 SMSs per day while a handful texts as high as 80 SMSs. This divide is because some people like texting and they subscribe to different SMS-friendly subscriptions.

[7] One can install it by downloading from http://www.vacxq.com/ActivPass/



(a) Histogram of average #calls per day across participants

(b) Histogram of average #SMS per day across participants

Figure 12. Phone activity related statistics across participants

EXPERIMENTAL RESULTS

In this section, we present the detailed evaluation results of our system. We conduct the evaluation process in two different ways. First we refine the generation of question set based on the lessons learnt during the pilot study (discarded yes/no questions, questions are generated from last three days activities, title is provided instead of web URL etc.) and then perform an extensive evaluation to decide number of questions needed in final challenge. Hence the important difference in system between pilot study and second stage lies in question selection where the pilot study essentially draws questions from a superset. In this regard, we recruit volunteers and evaluate the quality of the new set of generated questions both in terms of recall rate and guessability. Based upon the feedback from the evaluation process, we go ahead to build an end-to-end system and compare our system with several baseline methods. We have also conducted a user survey to assess the usability of our system.

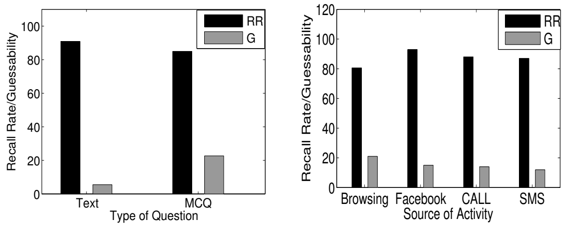
Evaluation -Second Stage

Each of the 50 recruited participants (22 under-graduate, 9 graduate students, 11 research scientist, 2 lab-assistants, 4 faculty members, and 2 academic-staffs) has been requested to identify three to five friends among the participants. The generated question set consists of 35 questions, 20 questions pertain to participant's own activities and 15 pertain to participant's *friends'* activities (of previous two days). We obtain 1750 responses in total among which 1000 responses were about user's own activity and 750 responses were about *friends'* guess. Several sources of information and question format have been discarded after the first round. Questions were uniformly generated from all remaining question-format and sources.

Observations

We report the improvements that we achieve as a result of the refinement process. The improvement is multidimensional; we observe the improvement due to (a). refinement in the question formats, (b). carefully selecting the sources, (c). restricting to last 2 days activities significantly improves the performance.

(a) Question Format-wise Recall Rate and Guessability: First we highlight the results in Fig. 13(a) which shows the substantial improvement in recall rate and guessability across different question formats. As we have discarded yes/no format questions during the design phase, we are left with only MCQ and text based questions. Recall rates of MCQ and text based questions are 85% and 90.9% respectively while their respective guessabilities are 22.7% and 5.7% respectively. The recall in text-based question is low because of its low recall in web page category; text-based password is not tried in this category because it becomes too lengthy.



(a)

(b)

Figure 13. Fig. 13(a): Recall Rate and Guessability across different question-format. Fig. 13(b): Recall Rate and Guessability of activities from different sources

(b) Source-wise Recall Rate and Guessability: Similarly, Fig. 13(b) manifests the improvements that we achieve in the source-wise recall and guessability rate. Improvement of questions increases source-wise recall rate significantly.

On the other hand guessability rate of all sources reduces but the guessability of text-based passwords increase a bit because at this phase we are using a two-character hint universally. On dissecting every source result question format-wise (Table 6), we find that guessability of text-based questions is significantly lower, while recall is also higher than MCQ.

Table 6. Recall Rate and Guessability for various source and question-format combinations

Question-format Source	MCQ	Text-based
Facebook	93.9, 25.4	92.3, 6.5
Web	80.6, 21	N/A, N/A
Call	87.5, 23.6	90, 5.9
SMS	88, 20.8	86.2, 4.5

User-wise Recall and Guessability: Fig. 14 shows user-wise recall and guessability of all 50 users. Recall rate varies from 60% to 100% while mean and standard deviation of recall are 86.3% and 9.5 respectively. Guessability of user varies from 3.2% to 27.8% while mean and standard deviation of guessability are 14.6% and 5.75 respectively. We

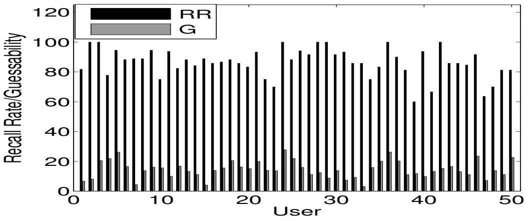


Figure 14. Recall Rate and Guessability of all the 50 users

find that for each user, recall rate is substantially higher than guessability.

End-to-End $ActivPass$ System

We are now in a stage of transforming this *refined system* to a final end-to-end $ActivPass$ system. Unlike user-study, a working system cannot afford to ask 20 questions. Hence a practical setting would be to generate a small challenge set which would have n questions and a user would be authenticated if she answers k questions correctly. We can derive the probability of success and failure for a given (n, k) pair by using a simple Bayesian formula. The table 7 shows the probabilities at different values of n and k.

Table 7. Success of authentic user and impostor with different n and k values

n	k	Authentic user	Impostor
4	4	0.554	0.0004
4	3	0.906	0.011
4	2	0.989	0.1043
4	1	0.998	0.468
3	3	0.642	0.0031
3	2	0.948	0.0577
3	1	0.996	0.3771
2	2	0.745	0.0213
2	1	0.981	0.2707

From the table we observe that a good value can be n = 3, k = 2 as it keeps the number of questions to bare minimum as well provides a healthy prediction of high likelihood of proper identification of authentic user and rejection of impostors.

Evaluation of $ActivPass$ System

In this section, we perform the performance evaluation of the final $ActivPass$ System. The performance is measured based on the two baseline schemes namely (a) ActivPass Random baseline (b) Keystroke Baseline, which will be described shortly. For the evaluation, we recruited 15 volunteers[8]. $ActivPass$ generates a final challenge set consisting of 3 questions. Every participant is asked to take the test 10 times - 4 times the challenge set is generated from their own activity while 6 times from their friends activity.

Baseline Schemes

To the best of our knowledge, a system to stop shareability is not in place. Hence, no straightforward instance of baseline scheme is available. So we define the following two baseline schemes.

(a) Keystroke Baseline: We consider a scheme which restricts password sharing by matching key stroke pattern [14]. We replicated this work to compare with $ActivPass$ as benchmark. This benchmark will be referred as *keystroke*. The basic intuition is that a service in the process of authenticating a person may not only check the password but also the pattern (key pressing duration, inter key press duration) in which the password is entered through keyboard.

(b) $ActivPass$ **Random Baseline:** We propose another baseline scheme similar to $ActivPass$ where instead of outlier activities, all activities are given equal importance during generating challenges. This benchmark will be referred as $ActivPass$ *random*.

Evaluation Results

We start the evaluation of the $ActivPass$ system by considering it as a stand alone system and measuring its performance independently. Table 8 shows user-wise *success* and *failure* of $ActivPass$ system. It shows that 12 out of 15 participants succeeded in answering the 4 challenges given to them while 2 users succeeded in 3 out of 4 challenges and 1 user got success in 2 out of 4 challenges. In case of challenges related to friends 11 users failed in all challenges while 3 users passed in 1 out of 6 challenges taken while 1 user passed in 2.

Table 8. Success and Failure rates for all the users

User ID	Success	Failure
1	1.0	0.0
2	1.0	0.0
3	1.0	0.166
4	0.75	0.0
5	1.0	0.0
6	1.0	0.166
7	1.0	0.0
8	0.5	0.0
9	1.0	0.33
10	1.0	0.0
11	0.75	0.0
12	1.0	0.0
13	1.0	0.166
14	1.0	0.0
15	1.0	0.0

Table 9. Comparison of ActivPass with benchmark schemes

Scheme	Success	Fail
ActivPass	0.95	0.055
ActivPass Random	0.95	0.35
Keystroke	0.744	0.397

Comparing $ActivPass$ **with Baseline schemes:** Table 9 highlights the superiority of $ActivPass$ system over the other benchmark schemes. It shows that $ActivPass$ system is far better than $keystroke$ method with respect to both *success* (fraction of attempt an authentic user successfully logged in) and *failure* (fraction of attempt an impostor successfully logged in). Success of $ActivPass - Random$ scheme is same as $ActivPass$, which means people equally remember usual activities as outlier, however, guessability of usual activities is much too high compared to outlier activities. Here we must mention that the system can become even more secure if we completely switch to text-based question.

User Survey on Ease of usage of ActivPass

We created a $Google$ form with a set of questions to gather feedback from users about their experience with $ActivPass$. Any password system need to be convenient for users specially it should be easy to adopt and use. Fig. 15(a) and

[8] 5 and 10 participants from pilot study and second stage respectively

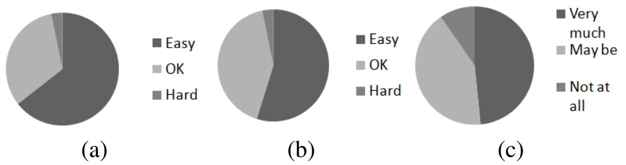


Figure 15. Users' Opinion - 15(a): Easy to adopt. Fig. 15(b): Effort required to remember their activities. Fig. 15(c): People's preference of ActivPass instead of hint question in password recovery

Fig. 15(b) show the users' opinion regarding adopting this new scheme. Fig. 15(c) shows the fraction of users eager to use the new scheme instead of traditional hint questions for password recovery. The results in general are encouraging.

CONCLUSION

This paper presents *ActivPass*, a dynamic authentication system that mines the user's daily activities to extract passwords. While ActivPass may not apply to services requiring strict authentication, it may be a candidate to alleviating the problem of password sharing. Even though users might share their passwords once, they are generally unwilling to continuously share their daily (atypical) activities with others. This can prevent Bob from perennially reusing Alice's (Netflix) password, just because she shared the password once. Experiment results from a large set of university volunteers demonstrate promising results with the system achieving up to 95% success rate. We also observe that while performing our experiments, volunteers were not penalized for failing to recall their past activities. In reality, however, a user has a stronger incentive to recall her past to be able to answer the password question correctly – in such situations, the performance could improve further. Of course, the adversary could also have a stronger incentive in reality to guess a user's password. Systematically understanding the actual performance in truly real-world situations is left to future work.

ACKNOWLEDGEMENTS

We want to convey our gratitude to ITRA (project code:DISARM-NUP), DST (project code: BCF), IIMA (project code: DNASUDA) for partially funding this work. We also want to convey our thanks to Prof. Shamik Sural and Dr. Sourangshu Bhattacharya of IIT Kharagpur and anonymous reviewers for their invaluable suggestions.

REFERENCES

1. Akhilomen, J. Data Mining Application for Cyber Credit-card Fraud Detection System. In *ICDM* (2013), 218–228.
2. Blonder, G. E. Graphical Password, U.S. Patent 5559961. `http://www.freepatentsonline.com/5559961.html`, Sept. 1996.
3. Bonneau, J., Herley, C., Oorschot, P. C. v., and Stajano, F. The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes. In *IEEE S & P* (2012), 553–567.
4. Chibelushi, C., Deravi, F., and Mason, J. A Review of Speech-Based Bimodal Recognition. *IEEE Transactions on Multimedia 4*, 1 (2002), 23–37.
5. Darwish, A. A., Zaki, W. M., Saad, O. M., Nassar, N. M., and Schaefer, G. Human Authentication Using Face and Fingerprint Biometrics. In *CICSyN* (2010), 274–278.
6. Das, S., Hayashi, E., and Hong, J. I. Exploring Capturable Everyday Memory for Autobiographical Authentication. In *UbiComp* (2013), 211–220.
7. de Vel, O., Anderson, A., Corney, M., and Mohay, G. Mining e-Mail Content for Author Identification Forensics. *SIGMOD Rec. 30*, 4 (2001), 55–64.
8. Denning, D. E., and MacDoran, P. F. Internet besieged. ACM Press/Addison-Wesley Publishing Co., 1998, ch. Location-Based Authentication: Grounding Cyberspace for Better Security, 167–174.
9. Gafurov, D., Helkala, K., and Sondrol, T. Biometric Gait Authentication Using Accelerometer Sensor. *JCP 1*, 7 (2006), 51–59.
10. Granger, E., Khreich, W., Sabourin, R., and Gorodnichy, D. O. Fusion of Biometric Systems Using Boolean Combination: An Application to Iris-Based Authentication. *Int. J. Biometrics 4*, 3 (2012), 291–315.
11. Guha, S., Jain, M., and Padmanabhan, V. Koi: A Location-Privacy Platform for Smartphone Apps. In *NSDI* (2012), 183–196.
12. Gupta, P., Wee, T. K., Ramasubbu, N., Lo, D., Gao, D., and Balan, R. HuMan: Creating Memorable Fingerprints of Mobile Users. In *PerCom* (2012), 479–482.
13. Khan, M. K., Zhang, J., and Wang, X. Chaotic Hash-Based Fingerprint Biometric Remote User Authentication Scheme on Mobile Devices. *Chaos, Solitons & Fractals 35*, 3 (2008), 519–524.
14. Mandujano, S., and Soto, R. Deterring Password Sharing: User Authentication via Fuzzy C-Means Clustering Applied to Keystroke Biometric Data. In *ENC* (2004), 181–187.
15. McCool, C., et al. Bi-Modal Person Recognition on a Mobile Phone: Using Mobile Phone Data. In *ICMEW* (2012), 635–640.
16. Nosseir, A., Connor, R., Revie, C., and Terzis, S. Question-Based Authentication Using Context Data. In *NordiCHI* (2006), 429–432.
17. Peacock, A., Ke, X., and Wilkerson, M. Typing Patterns: A Key to User Identification. *IEEE S & P 2*, 5 (2004), 40–47.
18. Yampolskiy, R. Human Computer Interaction Based Intrusion Detection. In *ITNG* (2007), 837–842.
19. Zorkadis, V., and Donos, P. On biometrics-Based Authentication and Identification from a Privacy-Protection Perspective: Deriving Privacy-Enhancing Requirements. *Inf. Manag. Comput. Security 12*, 1 (2004), 125–137.