مقدمه

رقم سمت راست شماره دانشجویی شما با شرکت/پژوهشکده/ هلدینگ‌هایی که در ادامه ذکر شدند تطابق دارد که برای پاسخ به سایر سوالات به آن نیاز دارید؛ اگرچه می‌توانید به دلخواه خود به بررسی یک سیستم دیگر بپردازید اما نباید توسط دانشجوی دیگری انتخاب شده باشد لذا می‌توانید نام شرکت/پژوهشکده/ هلدینگ مورد نظر خود را همراه شماره دانشجویی خود به دستیار آموزش ایمیل نمایید و پس از تأیید نهایی به سوالات پاسخ دهید.

0 : شرکت شاتل	1 : شرکت دیجی‌کالا	2 : شرکت دیوار
3 : پژوهشکده بیمه	4 : شرکت کافه بازار	5 : شرکت علی بابا
6 : شرکت ابرآروان	7 : پژوهشکده رویان	8 : شرکت نماوا
9 : شرکت اسنپ		

سایر شرکت‌های پیشنهادی:

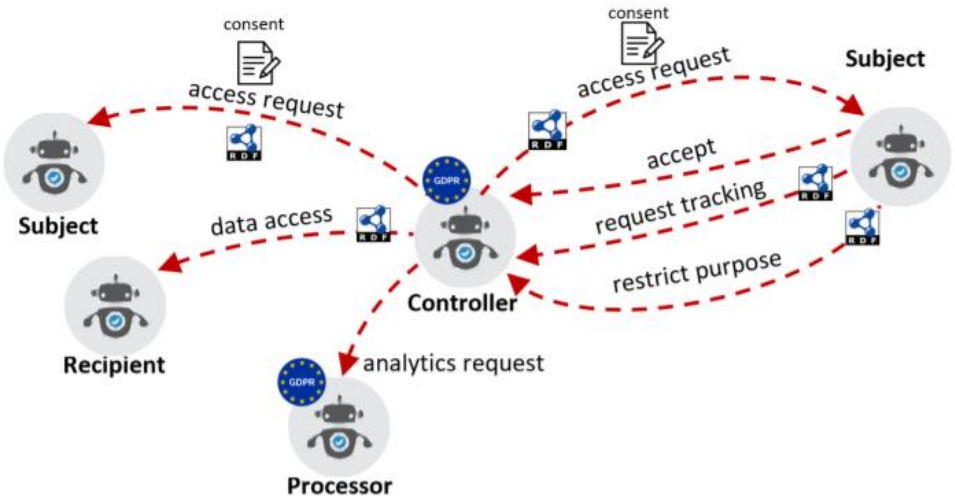
• نت برگ	کارگزاری مفید	همکاران سیستم
• گلرنگ سیستم	آپارات	سحاب
• هلدینگ آفرینش	. . .	

بخش اول – GDPR principles

1- محورهای اصلی محافظت و نگهداری از داده مطابق جلسات درسی طبق این لینک[1] معرفی شدند، در ابتدا اصـــول GDPR را نام برده و مزایای GDPR را با روش‌های محافظت از داده که تا ســـال 1998 استفاده می‌شدند و به (the 1998 Act) معروف هستند را در یک جدول همانند جدول زیر مقایسه کنید.

	the 1998 Act	GDPR
. . .		

2- چالش‌های رعایت قوانین حقوقی (قابل رفع و غیر قابل) را مطابق ســیســـتم تطبیق داده شـــده با شماره دانشجویی خود، با مشخص کردن عامل‌ها (Agents) و روابط بین آن‌ها همانند تصویر زیر با اصول GDPR تطبیق داده، اهمیت، تقدم و تأخّر هر یک از اصول GDPR را در این روابط شرح دهید. (هدف این ســوال آشـــنایی با عامل‌های مختلف یک ســیســـتم، روابط بین آن‌ها و مفاهیم processor و controller است)

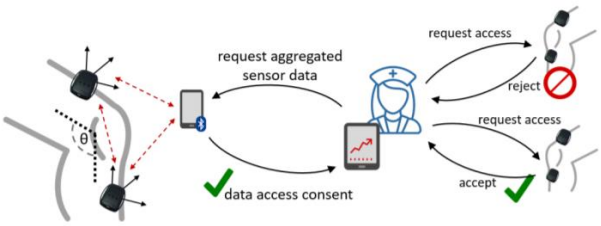


[1] https://gdpr-info.eu/

بخش دوم – Fairness biases

با توجه به عدم قطعیت موجود بین عامل‌های (Agents) یک محیط در رعایت عدالت (justice) بین آن‌ها و تفاوت مابین مفاهیم عدم قطعیت (uncertainty) و ابهام (ambiguity)، به کمک سیستم تطبیق داده شده با شماره دانشجویی خود، یکی از اهداف زیر را برگزینید؛ از آنجایی که در این محیط بیش از یک عامل حضور دارد، پس عامل‌ها می‌توانند رویکرد رقابتی (Competitive) و یا همکارانه (Cooperative) بگیرند. (برای مطالعه بیشتر می‌توانید به این لینک[1] مراجعه کنید)

- تابع هدف:کاهش تعداد درخواست‌های رد شده (rejected requests)
- تابع هدف: کمینه کردن فاصله «نرخ حداقل میزان رضایت» از «نرخ حداکثر میزان نارضایتی»
- تابع هدف: کاهش زمان تعامل بین عامل‌ها

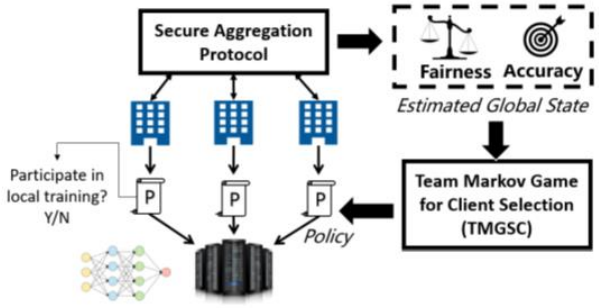


3- چه تصمیم‌هایی در سیستم شما باعث ایجاد bias در تقابل/همبستگی با یکدیگر می‌شود؟

4- با افزودن چه عبارتی (term) به تابع هزینه می‌توانیم به طور همزمان دقت (accuracy) و انصاف (fairness) را افزایش دهیم؟ با توجه به سیستم خود آیا این روش بهترین راه برقراری trade-off بین دقت و انصاف است؟ در صورت امکان به سایر روش‌ها اشاره کنید. **(امتیازی)**

5- در چه صورتی می‌توانیم هر 3 هدف معرفی شده را در یک مسئله راضی (satisfy) کنیم؟

6- مطابق شکل زیر یک چهارچوب مناسبی جهت یادگیری تابع هدف نهایی (حاوی تمامی اهداف) در محیط دارای ابهام به عامل‌ها ارائه دهید که زمان یادگیری آن‌ها را کمینه کند.



[1] https://mrtz.org/nips17

بخش سوم –Anonymization techniques

7- با توجه به ویژگی‌های موجود در پایگاه داده سیستم تخصیص داده شده به شماره دانشجویی خود، روش‌های مختلف بی‌نام سازی(anonymization) داده‌ها را روی حداقل 10 ستون بررسی کنید. (برای مثال زمان خرید، عرض جغرافیایی، کدپستی،شماره کارت اعتباری و ...)

(دقت شود که لزوماً روش حذف روش خوبی نیست؛ مزایا و معایب روش‌های مختلف کلاسیک را پیدا کرده و بررسی نمایید)

- Delete , The count-tree , Generalization
- Pseudonymization , Direct anonymization
- Suppression / Wiping
- . . .

8- در حوزه استخدام منابع انسانی (HR) می توان از روش های Anonymization استفاده کرد؟

9- داده‌ها پس از استفاده از روش Pseudonymization قابل بازیابی هستند؟ پیشنهاد شما چیست؟

10- پس از مشاهده شرکت Syntho[1]؛ مزایا، معایب و ناتوانی روش‌های کلاسیک در بی‌نام سازی داده‌های سیستم خود را بررسی کرده و روشی جایگزین پیشنهاد دهید.

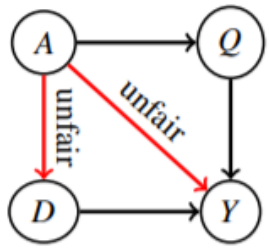
11- بروی یک مدل ساده شده از سیستم خود سه معیار اصلی عدم تبعیض (non-discrimination) را بررسی کنید.

12- تحت چه شرایطی می‌توان هریک از روش‌های زیر را روی مدل ساده‌شده استفاده‌کرد؟ (امتیازی)

(شروط کافی برای استفاده از هر تکنیک زیر با معیارهای [لازم] عدم تبعیض تفاوت دارد)

- α-MON , z-Anonymity
- k-anonymity , De-anonymization , T-closeness , L-diversity
- Causality Based Approaches , Set-based Anonymization

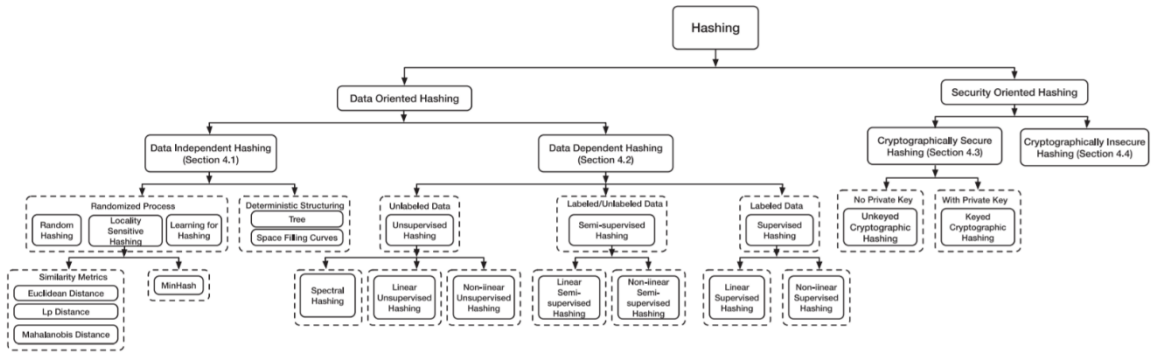
(می‌توانید دیاگرامی همانند شکل زیر برای مدل ساده شده خود رسم کنید)



[1] https://www.syntho.ai/

بخش چهارم – Hashing , reliability , redundancy

در سیستم‌های پیچیده (Complex Networks) انواع مختلفی از داده، روش‌های مختلفی برای hashing و به تبع آن با مشکلات متفاوتی روبرو خواهیم شد. در تصویر زیر یک طبقه‌بندی از تکنیک‌های hashing به صورت سلسله‌مراتبی با دو زیر گروه داده‌محور و امنیت‌محور مشاهده می‌کنید.



Cryptographic Hashing	**Keyed Cryptographic Hashing**	VMAC; UMAC; PMAC; OMAC; HMAC
		Poly1305-AES; MD6; BLAKE2
	Unkeyed Cryptographic Hashing	MD2/4/5/6
		SHA-1/3/224/256/384/512
		HAVAL; GOST; FSB; JH; ECOH
		RIPEMD/-128/-160/-320

13- با توجه به سیستم تخصیص داده شده به شماره دانشجویی خود، روش‌های مناسب برای hash کردن انواع مختلف داده‌هایی که در اختیار دارید (موقعیت‌مکانی، تصویر،صدا، متن و ...) را نام برده و دلیل استفاده از آن روش hashing را به اختصار توضیح دهید.

14- تفاوت بین دو زیر گروه امنیت‌محور و داده‌محور را به اختصار ذکر کنید.

15- در این تمرین با زیر گروه امنیت‌محور (به عنوان مثال روش های SHA و MD) از سلسله مراتب معرفی شده در تصویر بالا سروکار داریم؛ تابع hash در زیرگروه امنیت محور همانند روش‌های داده‌محور صرفاً یک ورودی با طول دلخواه را به یک خروجی با طول ثابت فشرده نمی‌کند؛ بلکه سه هدف (مقاومت در برابر برخورد، مقاومت در برابر preimage اول و دوم) را دنبال می‌کند در نتیجه از مشکلات عمده آن‌ها می‌توان به برخورد (collision) و brute force decrypt اشاره کرد برای رفع هر کدام از این مشکلات راه‌کاری‌های مختلفی ارائه دهید.

نکات تحویل

- مهلت ارسال این تمرین تا پایان روز جمعه 8 بهمن ماه خواهد بود.
- انجام این تمرین به صورت یک نفره می‌باشد.
- لطفا هر گونه فرض در حل سوالات را در گزارش خود ذکر کنید.
- لطفا گزارش ، فایل کدها و سایر ضمائم مورد نیاز را با فرمت زیر در سامانه مدیریت دروس بارگذاری نمائید.

HW8_[Lastname]_[StudentNumber].zip

- در صورت وجود سوال و یا ابهام می‌توانید از طریق رایانامه زیر با دستیار آموزشی در ارتباط باشید:

mohammad.nili@ut.ac.ir