



Data Centric Security and Data Protection

Manuela Cianfrone

Bologna 29/10/2016

Speaker

○ Manuela Cianfrone

EMEA Solution Architect @ Protegrity USA

- Implement Data Centric Security
- Design Data Security Solutions

Agenda

- Using Walls to Protect the Enterprise
- Data Centric Model
- Encryption & Tokenization
- De-Identification
- Use Cases



A Long Time Ago, in a Data Center Far, Far Away

Well, we have Jay, he
manages the firewall....



Walls

All kinds of walls...

- *physical walls*
- *access control*
- *DLP*
- *firewalls*
- *and many more*

Walls as Layers of Security

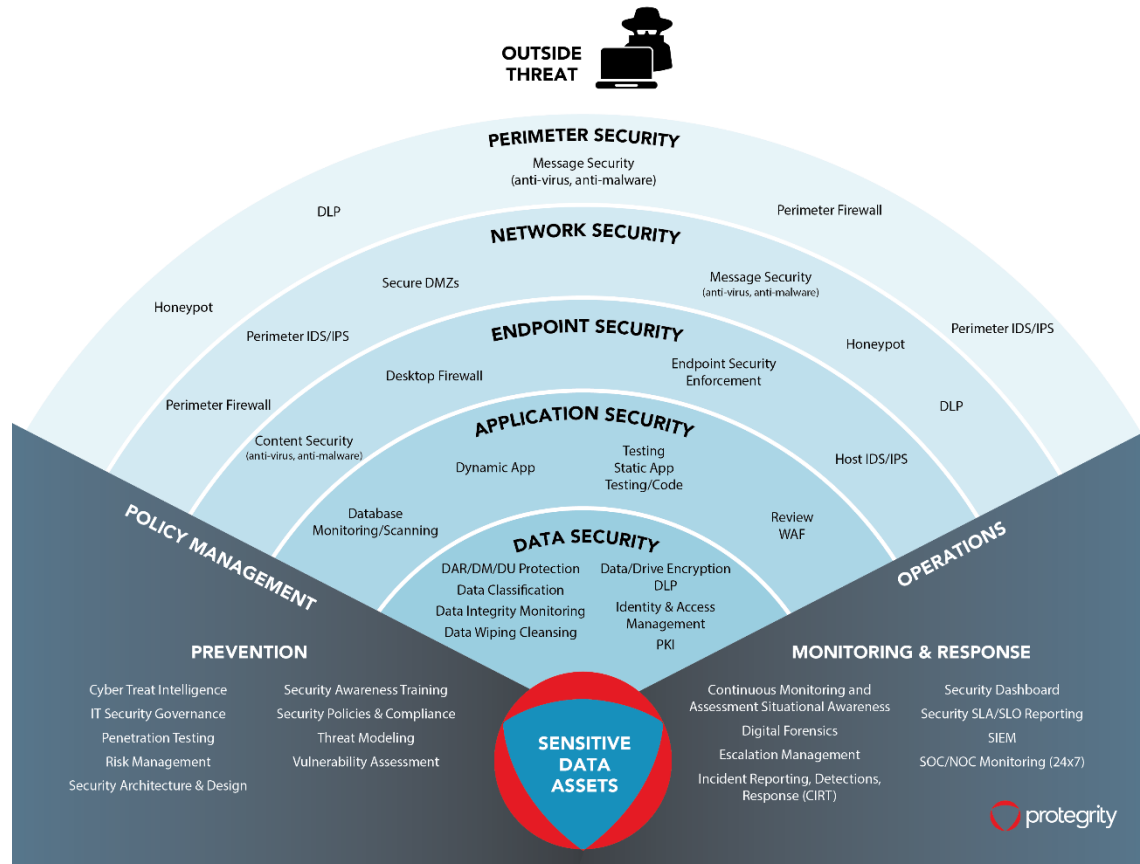
- Each wall of security provides an additional protection and control around your data.
- Each wall adds complexity.
- Each wall adds cost.
- Each wall adds overhead.
- Designs vary, all seeking a balance around securing the data versus impacting the users.



Walls Fail!!!

As evidence from the news headlines, insiders or hackers will get through the walls.

The Security Landscape – Focus on Data Centric Protection



Philosophy

The only way to secure sensitive data is to protect the data itself.

Gartner's View

- The exponential growth in data generation and usage is rendering current methods of data security governance obsolete, requiring significant changes in both architecture and solution approaches.
- Organizations lack coordination of data-centric security policies and management across their data silos, resulting in inconsistent data policy implementation and enforcement.
- Data cannot be constrained within storage silos but is constantly transposed by business processes across multiple structured and unstructured silos on-premises or in public clouds.

Data Centric Security

- Data Classification
- Data Discovery
- Centralized Security Policy Management
- Monitoring of User Privileges and Activity
- Auditing and Reporting
- Fine Grained Data Protection

Classification and Discovery

○ Data Classification Considerations

- Who should be able to access and maintain the data?
- What legal or regulatory requirements apply?
- What is the risk to the business if the data is compromised or disclosed?
- What is the data value?
- Where is the data stored?
 - Which systems, tables, columns, fields, files?

Classification and Discovery Complete

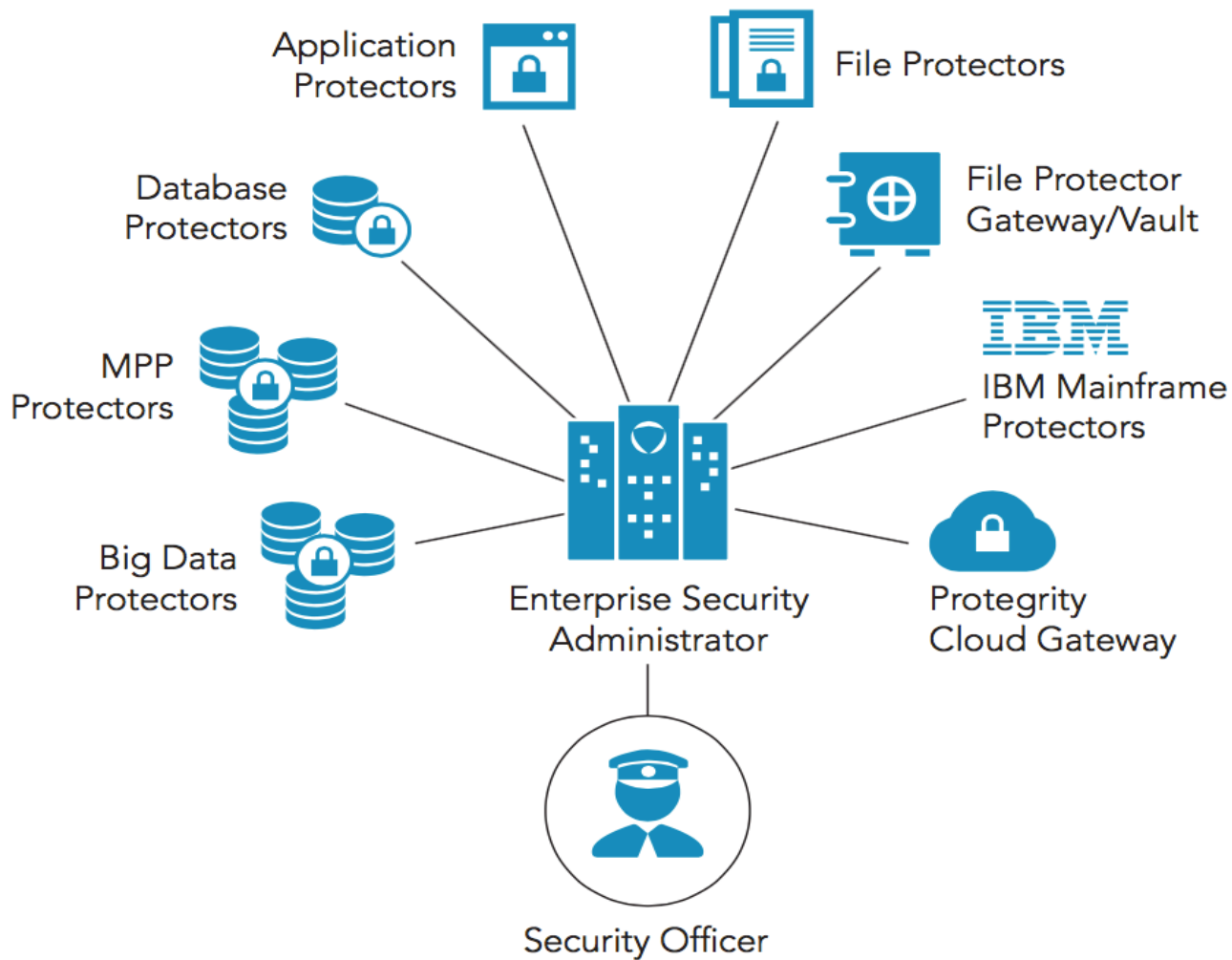
This is your Data Security Policy!

Identifier	What	How	Who	When	Where	Audit
Name	DE_NAME					
Address	DE_ADDRES S					
Date of Birth	DE_DOB	Monitor	HR, DS_Hadoop		EDW, Hadoop	Unauthorized Authorized
Social Security Number	DE_SSN	Tokenize All	HR		EDW, Hadoop	Unauthorized Authorized
Credit Card Number	DE_CCN	Tokenize (expose first 6, last 4)	Payments, CSR	9 – 5, M - F	EDW, Hadoop	Unauthorized Authorized
E-mail Address	DE_EMAIL	Tokenize All	HR, CSR, DS_Hadoop		EDW, Hadoop	Unauthorized Authorized
Telephone Number	DE_TELEPHO NE					

Centralized Policy Management

- Classify once, apply everywhere
 - Once classified, the data must be protected consistently.
 - Silo based approaches leave gaps in capability, management and controls.
 - A centralized policy applied to data across all silos is required.

Centrally Managed Cross Platform Policy Deployment



Monitoring, Auditing, Reporting

- Who has access to the data?
- When are they accessing the data?
- Where are they accessing the data?
- Why are they accessing the data?
- How are they accessing the data?
- Regular reporting, review and approval.
- Alerting on anomalous behavior.

Fine Grained Data Protection

- Provide access based on the least required for the use case
- Control access at the field level, or even within the field.
- Time based access control
- Segregate sensitive network, systems, applications and/or users whenever possible.
- De-Identify data when possible.

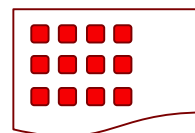
Granularity of Protecting Sensitive Data

Coarse Grained Protection (File/Volume)



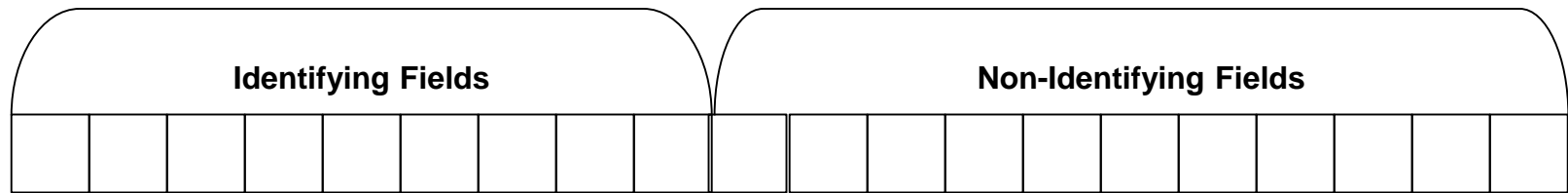
- Methods: File or Volume encryption
- “All or nothing” approach
- Does NOT secure file contents in use
- OS File System Encryption
- HDFS Encryption
- Secures data at rest and sometimes in transit

Fine Grained Protection (Data/Field)



- At the individual field level
- Fine Grained Protection Methods:
 - Vaultless Tokenization
 - Encryption (Strong, Format Preserving)
- Data is protected wherever it goes
- Business intelligence can be retained

Data Centric Security – Fine Grained Access Control



Identifying Fields

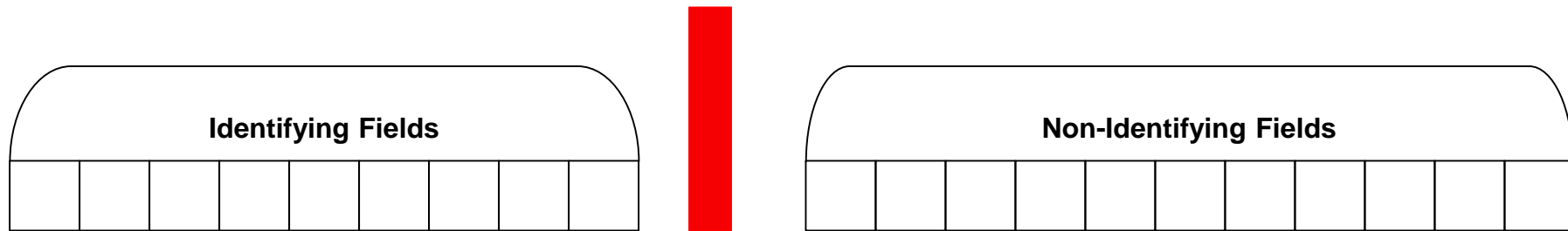
- First Name, Last Name
- Address
- Drivers License
- Social Security
- Date of Birth
- Credit Card Numbers
- Location
- Etc.

Non-Identifying Fields

- Salary
- Healthcare condition
- Account balances
- Account transaction details
- Etc.



De-Identified Information



The identifiable fields are de-coupled from the information about that individual.

The data on the individual cannot be associated with the individual.



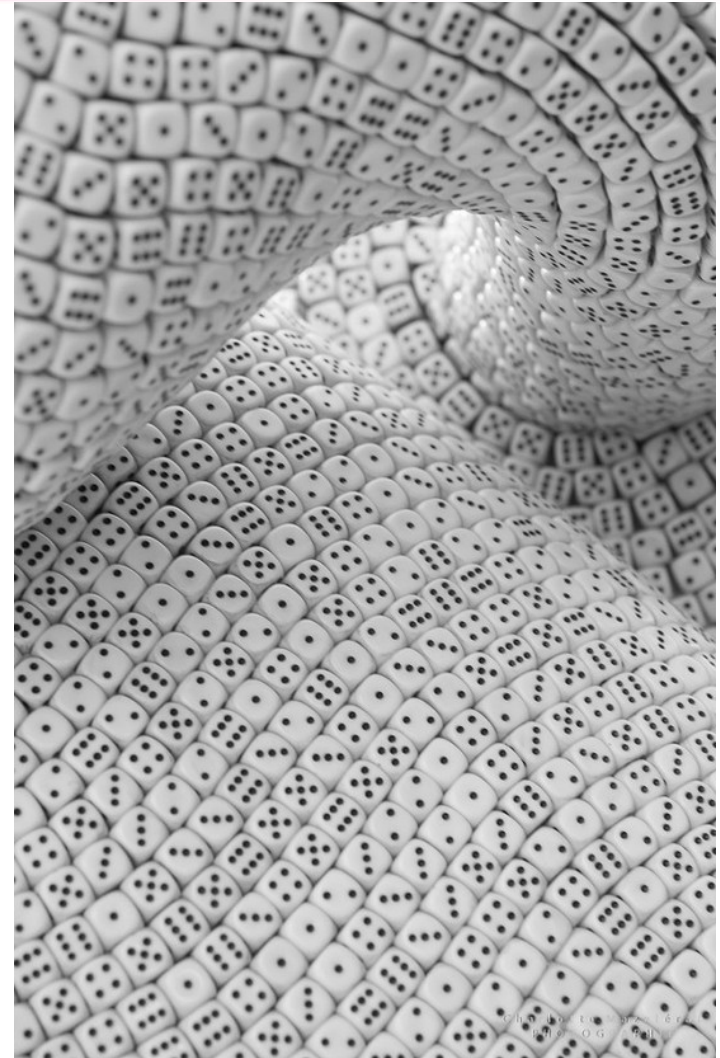
Using Encryption

- Encryption - A mathematically reversible cryptographic function, based on a known strong cryptographic algorithm and strong cryptographic key.
- Direct mathematical relationship between the plaintext, the ciphertext, the algorithm and the key.
- Ciphertext has minimal business value.
- Most usage requires access to sensitive data

p»|X!æF>½- {T%|ÊEe"\$c!³ aïf&±"ž
 B;k†, ?æcđ&DH%ÖFñ!ÖBp%\$X°UgT⁻
 ææ?9+|î?Üöö•AÔ₁æ)Îý;¿jÜø! 'Öx%
 ā&ā_T@JĀ¼- *|>-L|!!kÓnxZ6,, {F, æØQ'
 5| e¾@´ |p•ëp«%E´ _•JB(ðT|£.İ`>>
 ïr÷d~»±"üÛ™8´ ¯Ráõ±ÓO·eîKĀÎ
 *ötUĀīMÑz |H:%Sžžh?s ~V¾Ēø¶W
 rn;§"□_/Ā` ÷İ84Ò-Œb@α-¶TîâNS{
 y*³`àdđìš~|~|ªžk|*ùÿÈĒ Ü¿÷Ā™L

Using Tokenization

- Tokenization- Assignment through an index function, sequence number or a randomly generated number.
- No mathematical relationship between the data and the token.
- No algorithm, no key.
- A specific index must be referenced to connect the data and token.
- A Token is a non-sensitive replacement for sensitive data.
- Tokens have business value.
- Fewer users need sensitive data

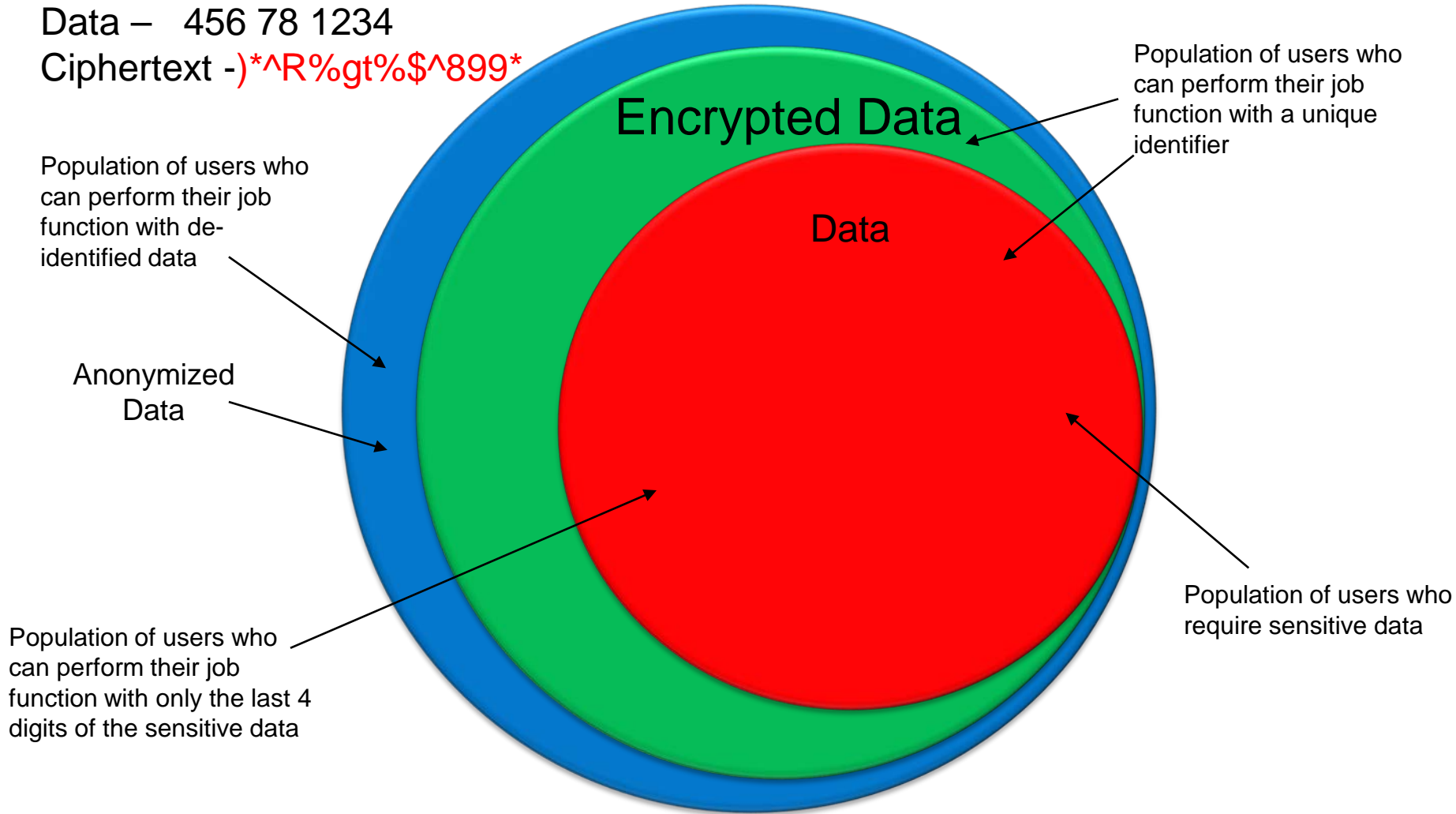




Encryption to Reduce Exposure and Risk

Data – 456 78 1234

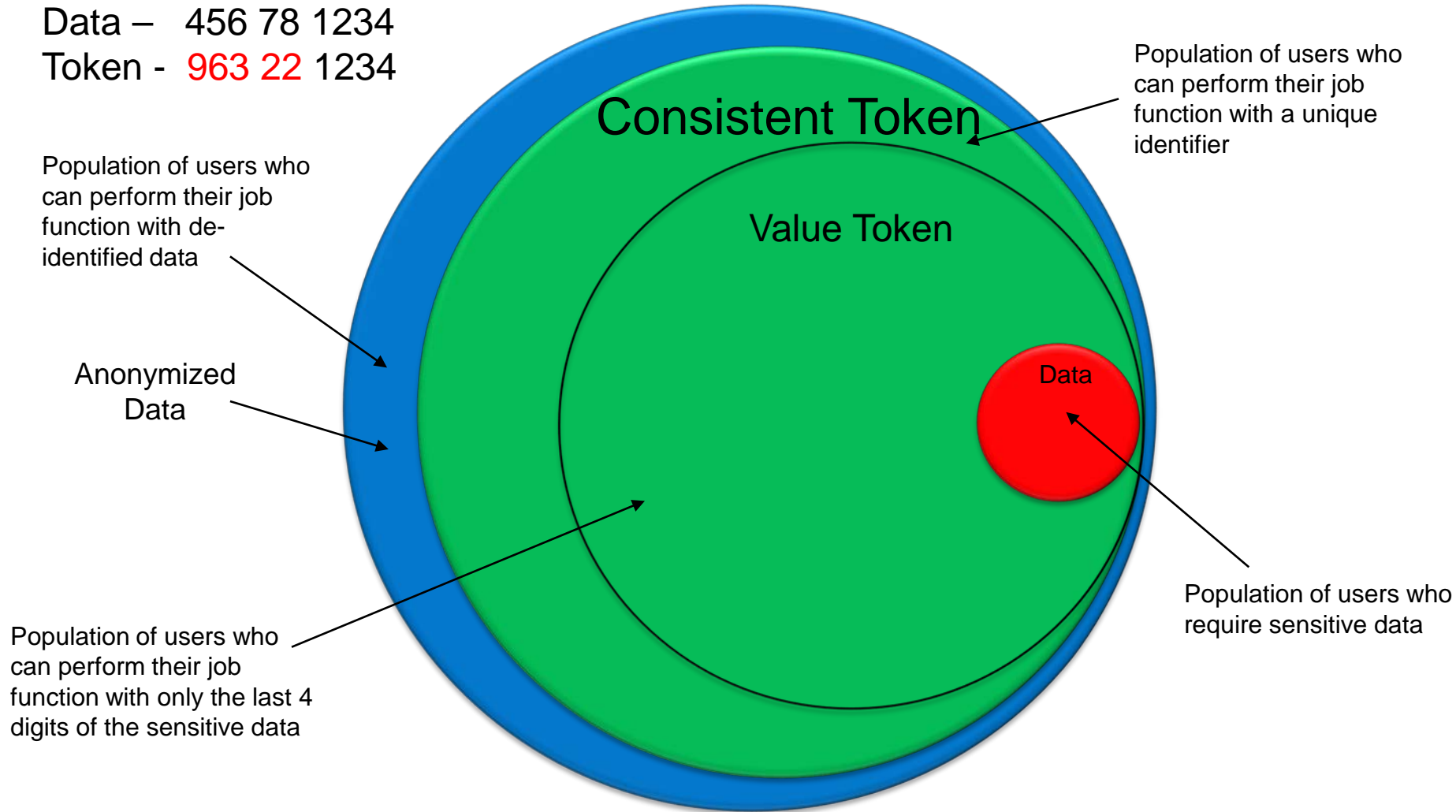
Ciphertext - `)*^R%gt%$^899*`






Tokenization to Reduce Exposure and Risk

Data – 456 78 1234

Token - 963 22 1234



De-Identified Sensitive Data

Field	Real Data	Protection: Tokenized
Name	Joe Smith	csu wusoj
Address	100 Main Street, Pleasantville, CA	476 srta coetse, cysieondusbak, CA
Date of Birth	12/25/1966	01/02/1966
Telephone	760-278-3389	760-389-2289
E-Mail Address	joe.smith@surferdude.org	eo.e.nwuer@beusorpdqo.org
SSN	076-39-2778	076-28-3390
CC Number	3678 2289 3907 3378	3846 2290 3371 3378
Business URL	www.surferdude.com	www.sheyinctao.com
Fingerprint		Encrypted
Photo		Encrypted
X-Ray		Encrypted
Healthcare / Financial Services	Identifiers such as name, address, email address, SSN, CCN, DoB, etc.	Healthcare Data, Spending Data, Financial Data

Comparing different de-identification approaches

- Methods of de-identifying PHI/PII include;
 - Suppression (Redaction)
 - Generalized Masking
 - Encryption (AES)
 - Pseudonymization - Vaultless Tokenization

Personally Identifiable Information / Protected Health Information							No need to protect!
Name	Address	Date of Birth	SSN	CCN	E-mail address	Telephone Number	Information about the individual
Joe Smith	100 Main Street, Pleasantville, CA	12/25/1966	076-39-2778	3678 2289 3907 3378	joe.smith@surferdude.org	760-278-3389	Financial Data Healthcare Data Spending data
xxx Smith	xxxxxxxxxxCA	xx/xx/1966	076xxxxxx	xxxxxxxxxxx3378	xxxxxxx@xxxxxxxx.org	760xxxxxx	
!@#%\$^a	!@#%\$^a^.,mhu7//&*B)_+!@	!@#%\$^a^.,mhu7///&	^.,mhu7///&*B)_+!@	!@#%\$^a^.,mhu7///&	!@#%\$^a^.,mhu7///&*B)_	#%\$^a^.,mhu7///&	
csu wusoj	476 srta coetse, cysieondusbak, HA	01/02/1983	478-389-0048	3846 2290 3371 3904	eo.e.nwuer@beusorpdqo.fol	478-389-2289	

Algorithm Properties

Properties ↓	Description
Strength	Known cryptographic analysis that will prove the strength of the algorithm. Typically in terms of the number of bits used by the key.
Where Used	Production or Non-Production. You want to be able to leverage your investment in both Production and Non-Production environments.
Performance	Highest performance for the algorithm used. This factor will be amplified with large number of security operations.
Transparency	Lowest level of changes that need to be made to the host business systems.
Reversibility	You would like to be able to deliver clear text data to authorized users.
Standards Based	You would like the algorithm to be supported by a standards body.
Usability and Analytics	You would like the algorithm to not place restrictions on analysis – not to hinder it.
Deployment Choices	In-process deployment is desirable to minimize performance degradation encountered with clustered deployment approach.
Applicability for PCI DSS	Is the algorithm usable for protecting credit cards under the PCI DSS guidelines?
Applicability for PII	Is the algorithm usable for protecting credit cards under the PII guidelines?
Applicability for PHI	Is the algorithm usable for protecting credit cards under the PHI guidelines? Particularly for HIPAA.

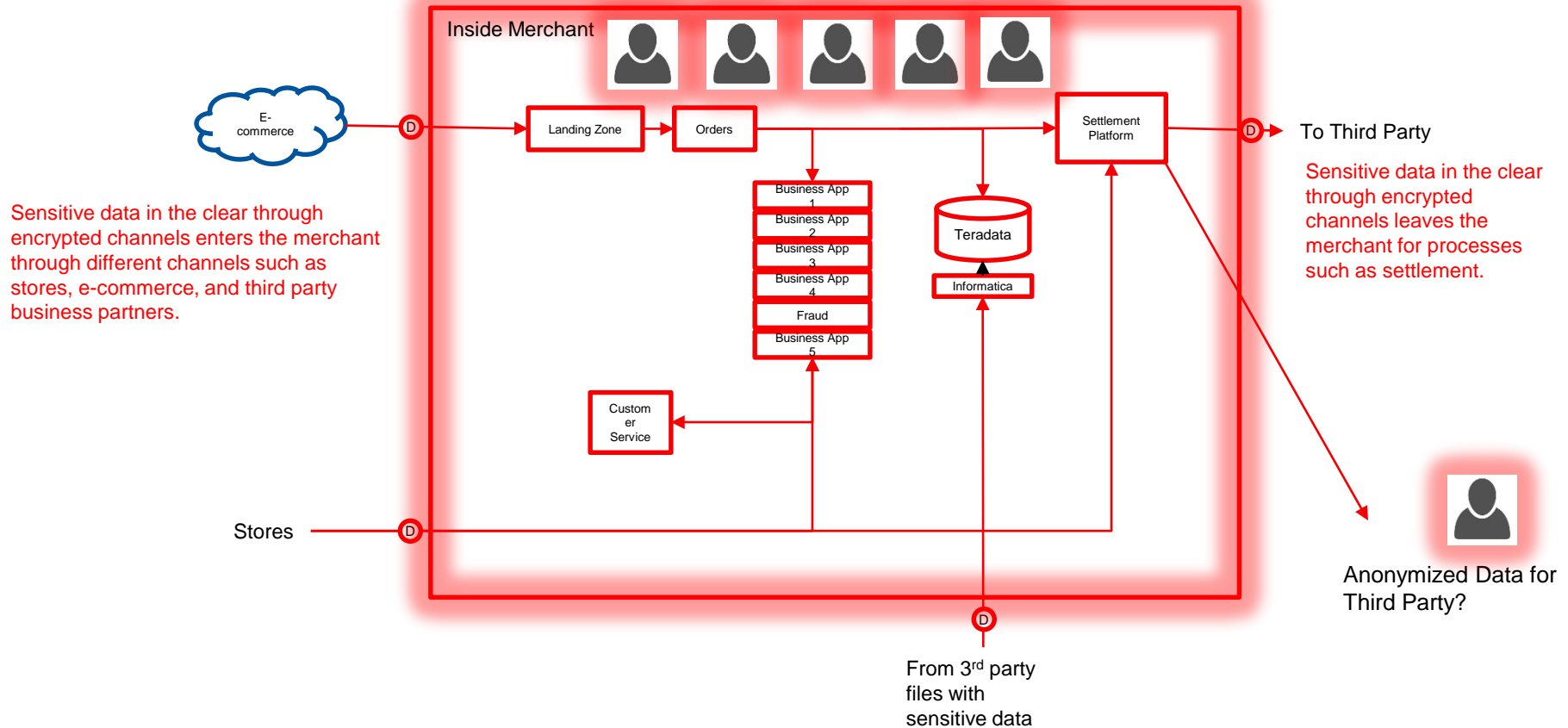
Algorithm

Algorithm →	Encryption (AES / TDES)	Vaultless Tokenization	Vault-based Tokenization	Format Preserving Encryption	Masking / Obfuscation
Properties ↓					
Strength	Strong	Strong	Strong	Strong	Strong - Medium
Where Used	Production	Production / Non-Production	Production / Non-Production	Production / Non-Production	Non-Production
Performance	Fastest	Fast	Slowest	Medium	Medium – N/A
Transparency	Poor	High	High	High	High
Reversibility	Reversible	Reversible	Reversible	Reversible	Not Reversible
Standards Based	NIST, FIPS & Others	None	None	None	None
Usability with Analytics	Medium	High	Medium	High	Medium
Deployment Choices	Cluster or In-Process	Cluster or In-Process	Cluster	Cluster or In-Process	N/A
Applicability for PCI DSS	Medium	Highest	High	Medium	Not Usable
Applicability for PII	High	Highest	Not Usable	High	Low
Applicability for PHI	High	Highest	Not Usable	High	Low

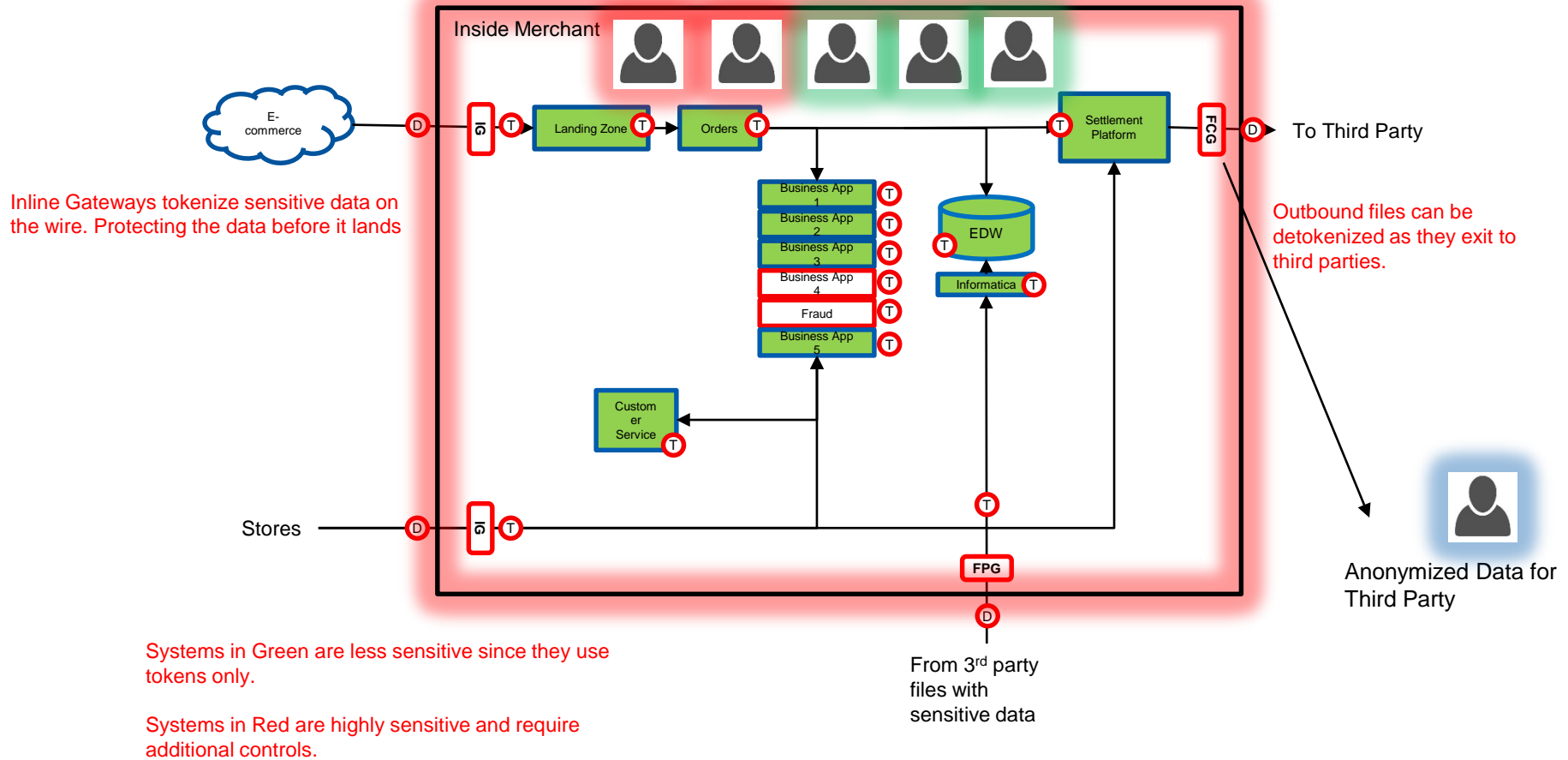
Algorithm

Algorithm →	Encryption (AES / TDES)	Vaultless Tokenization	Vault-based Tokenization	Format Preserving Encryption	Masking / Obfuscation
Properties ↓					
Strength	Strong	Strong	Strong	Strong	Strong - Medium
Where Used	Production	Production / Non-Production	Production / Non-Production	Production / Non-Production	Non-Production
Performance	Fastest	Fast	Slowest	Medium	Medium – N/A
Transparency	Poor	High	High	High	High
Reversibility	Reversible	Reversible	Reversible	Reversible	Not Reversible
Standards Based	NIST, FIPS & Others	None	None	None	None
Usability with Analytics	Medium	High	Medium	High	Medium
Deployment Choices	Cluster or In-Process	Cluster or In-Process	Cluster	Cluster or In-Process	N/A
Applicability for PCI DSS	Medium	Highest	High	Medium	Not Usable
Applicability for PII	High	Highest	Not Usable	High	Low
Applicability for PHI	High	Highest	Not Usable	High	Low
Totals	5	9	4	6	1

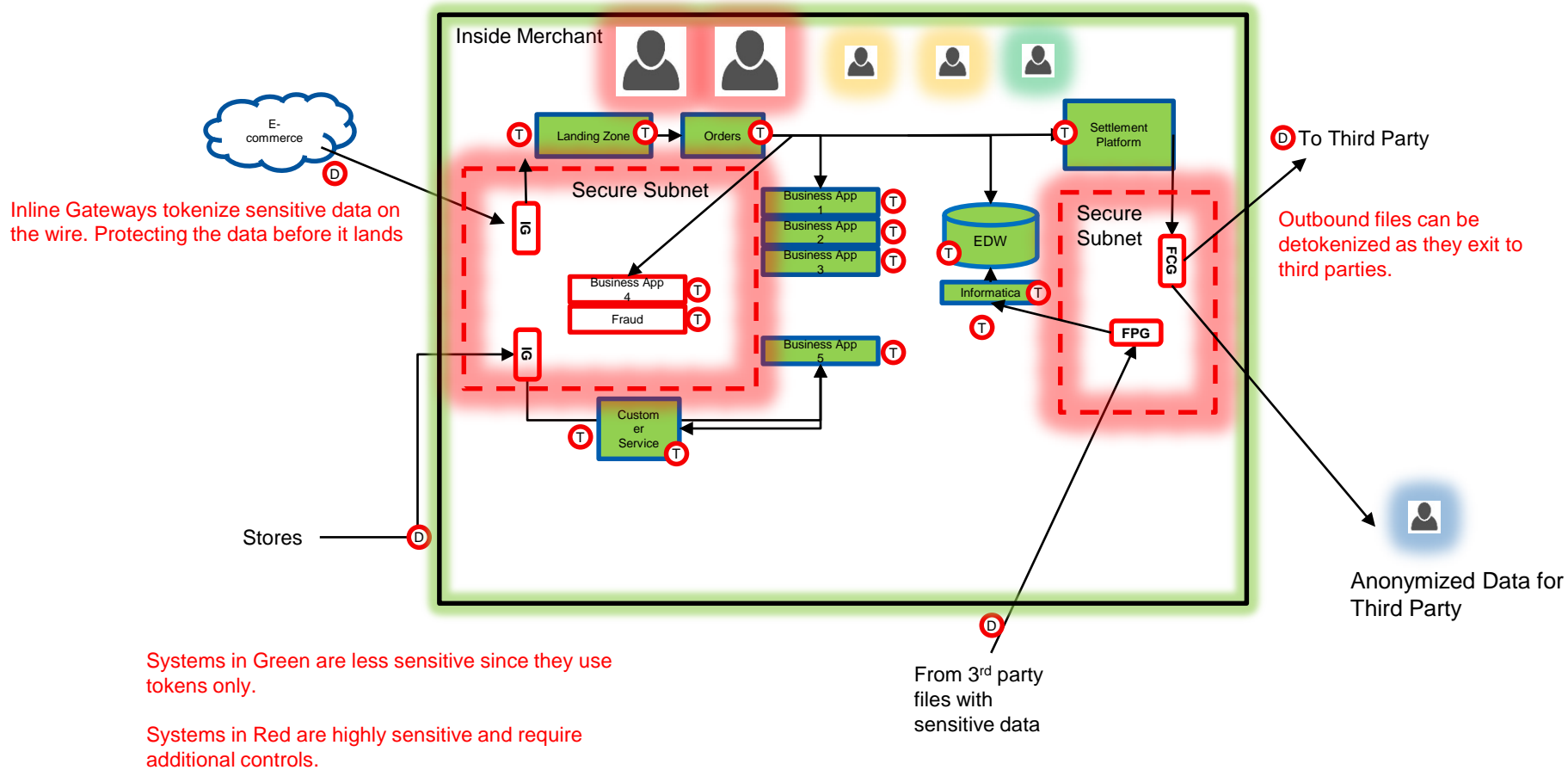
Sensitive Data Enters and Leaves Merchant Network



Data Centric Model – System Segmentation



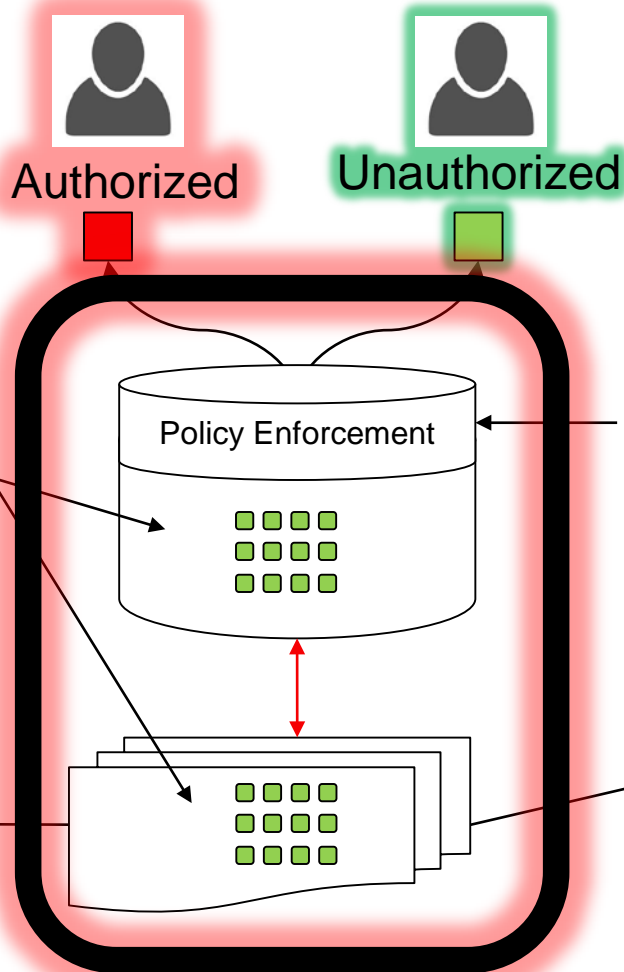
Data Centric Model – Network Segmentation



Data Centric Model – Database User Segmentation

Authorized Users:

Additional controls can be placed around Authorized Users (behavior monitoring/alerting, 2 FA etc.)



Unauthorized Users:

Can't get sensitive data in the clear, few controls required.

Fine Grained Data Protection

Is the best approach for protecting sensitive data.

Policy Enforcement for In Use Protection

Is used by security officers to authorize only the users who need to see sensitive data in the clear to perform their job duties. Keeps out the Privileged users.

Bad Guys:

Even if the bad guys get into the raw files that make up the database, they would be getting fake data.

Privileged Users:

Even if the privileged users get into the raw files that make up the database, they would be getting fake data.



Thank You